

# **Handbook of Risk Theory**



Sabine Roeser, Rafaela Hillerbrand, Per Sandin,  
Martin Peterson (Eds.)

# **Handbook of Risk Theory**

**Epistemology, Decision Theory,  
Ethics, and Social Implications of Risk**

With 85 Figures and 62 Tables

*Editors*

Sabine Roeser (Editor-in-Chief)  
Philosophy Department  
Faculty of Technology, Policy and Management  
Delft University of Technology  
Delft  
The Netherlands  
and  
Philosophy Department  
University of Twente  
Enschede  
The Netherlands

Rafaela Hillerbrand  
Human Technology Center  
RWTH Aachen University  
Aachen  
Germany

Per Sandin  
Department of Plant Physiology and Forest Genetics  
Swedish University of Agricultural Sciences  
Uppsala  
Sweden

Martin Peterson  
Section for Philosophy and Ethics  
Eindhoven University of Technology  
Eindhoven  
The Netherlands

ISBN 978-94-007-1432-8                    e-ISBN 978-94-007-1433-5  
Print and electronic bundle under ISBN 978-94-007-1434-2  
DOI 10.1007/978-94-007-1433-5  
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2011941145

© Springer Science+Business Media B.V. 2012

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

# Preface

Risk has become one of the main topics in fields as diverse as engineering, medicine, and economics, and it is also studied by social scientists, psychologists, and legal scholars. But the topic of risk also leads to more fundamental questions such as: What is risk? What can decision theory contribute to the analysis of risk? What does the human perception of risk mean for society? How should we judge whether a risk is morally acceptable or not? Over the last couple of decades, questions like these have attracted interest from philosophers and other scholars into risk theory.

This handbook provides an overview into key topics in a major new field of research. It addresses a wide range of topics, from decision theory, risk perception, to ethics and social implications of risk, and it also addresses specific case studies. It aims to promote communication and information among all those who are interested in theoretical issues concerning risk and uncertainty.

This handbook brings together leading philosophers and scholars from other disciplines who work on risk theory. The contributions are accessibly written and highly relevant to issues that are studied by risk scholars. We hope that the *Handbook of Risk Theory* will be a helpful starting point for all risk scholars who are interested in broadening and deepening their current perspectives.

The editors:

Sabine Roeser (Editor-in-Chief), Rafaela Hillerbrand, Per Sandin, and Martin Peterson



# Acknowledgments

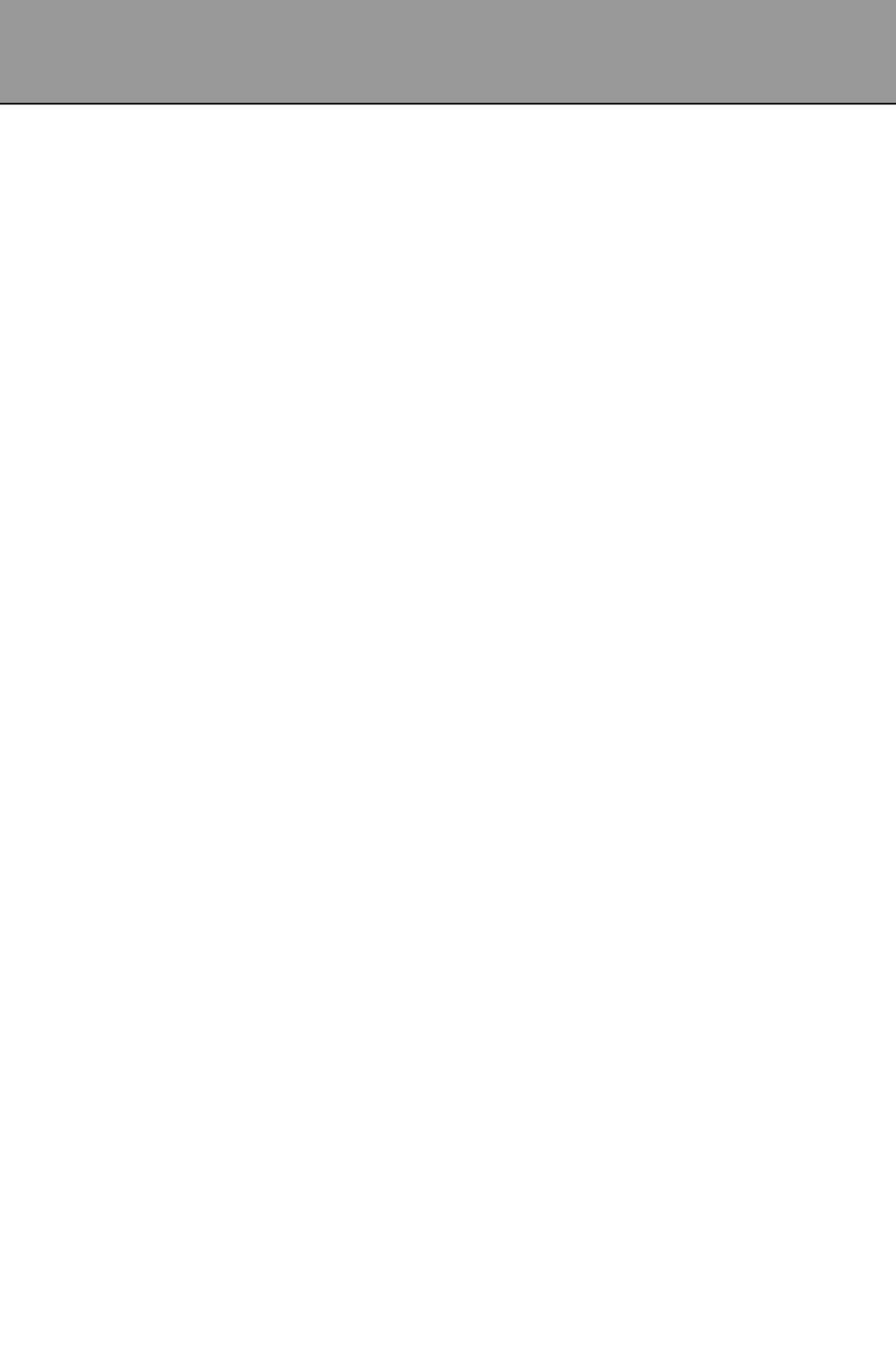
Sabine Roeser's work for the *Handbook of Risk Theory* has been conducted at the philosophy departments of TU Delft and Twente University and was sponsored by the Netherlands Organization for Scientific Research (NWO), with VIDI-grant number 276-20-012.

Rafaela Hillerbrand's work on this volume was supported by the excellence initiative of the German federal and state governments and conducted at the Human Technology Centre (HumTec) and the Institute of Philosophy, RWTH Aachen University. Thanks to all members of the research group eet - ethics for energy technology for insights into risks from various disciplines. Special thanks to Andreas Pfennig, Nick Shackel, and Peter Taylor for numerous fruitful discussions on the topic.

Per Sandin's work for this handbook was done in the Department of Plant Physiology and Forest Genetics, Swedish University of Agricultural Sciences, Uppsala and in the Department of Philosophy and History of Technology, Royal Institute of Technology, Stockholm.

Martin Peterson has conducted his work for the handbook at the philosophy department of TU Eindhoven. Sabine Roeser and Martin Peterson are members of the 3TU. Centre for Ethics and Technology, a center of excellence of the federation of the three technical universities in the Netherlands (Delft, Eindhoven, and Twente).

We are very grateful to the contributors of this handbook. In addition, we would like to thank Katie Steele, Linda Soneryd, and Misce Wester, each of whom provided expert reviews for a chapter. We would like to thank the staff at Springer for the excellent collaboration, specially Ties Nijssen, Jutta Jaeger-Hamers, and Christine Hausmann.



# Table of Contents

Preface .....	v
Acknowledgments .....	vii
Editors .....	xiii
List of Contributors .....	xv

## Volume 1

1 Introduction to Risk Theory .....	1
Sabine Roeser · Rafaela Hillerbrand · Per Sandin · Martin Peterson	

### Part 1 General Issues in Risk Theory

2 A Panorama of the Philosophy of Risk .....	27
Sven Ove Hansson	
3 The Concepts of Risk and Safety .....	55
Niklas Möller	
4 Levels of Uncertainty .....	87
Hauke Riesch	

### Part 2 Specific Risks

5 The Economics of Risk: A (Partial) Survey .....	113
Louis Eeckhoudt · Henri Loubergé	
6 Interpretation of Forensic Evidence .....	135
Reinoud D. Stoel · Marjan Sjerps	
7 Risks and Scientific Responsibilities in Nanotechnology .....	159
John Weckert	
8 Risk and Risk-Benefit Evaluations in Biomedical Research .....	179
Annette Rid	
9 Understanding and Governing Public Health Risks by Modeling .....	213
Erika Mansnerus	

<b>10</b>	<b>Management of the Risks of Transport</b>	<b>239</b>
	<i>John Adams</i>	
<b>11</b>	<b>Risk and Spatial Planning</b>	<b>265</b>
	<i>Claudia Basta</i>	
<b>12</b>	<b>Intergenerational Risks of Nuclear Energy</b>	<b>295</b>
	<i>Behnam Taebi</i>	
<b>13</b>	<b>Climate Change as Risk?</b>	<b>319</b>
	<i>Rafaela Hillerbrand</i>	
<b>14</b>	<b>Earthquakes and Volcanoes: Risk from Geophysical Hazards</b>	<b>341</b>
	<i>Amy Donovan</i>	

### **Part 3 Decision Theory and Risk**

<b>15</b>	<b>A Rational Approach to Risk? Bayesian Decision Theory</b>	<b>375</b>
	<i>Claus Beisbart</i>	
<b>16</b>	<b>A Philosophical Assessment of Decision Theory</b>	<b>405</b>
	<i>Karsten Klint Jensen</i>	
<b>17</b>	<b>The Mismeasure of Risk</b>	<b>441</b>
	<i>Peter R. Taylor</i>	
<b>18</b>	<b>Unreliable Probabilities, Paradoxes, and Epistemic Risks</b>	<b>477</b>
	<i>Nils-Eric Sahlin</i>	
<b>19</b>	<b>Paradoxes of Rational Choice Theory</b>	<b>499</b>
	<i>Till Grüne-Yanoff</i>	
<b>20</b>	<b>Multi-Attribute Approaches to Risk</b>	<b>517</b>
	<i>Paul Weirich</i>	
<b>21</b>	<b>Real-Life Decisions and Decision Theory</b>	<b>545</b>
	<i>John R. Welch</i>	
<b>22</b>	<b>Social Influences on Risk Attitudes: Applications in Economics</b>	<b>575</b>
	<i>Stefan T. Trautmann · Ferdinand M. Vieider</i>	

## Volume 2

### Part 4 Risk Perception

<b>23</b>	<b>Risk Intelligence</b>	<b>603</b>
	<i>Dylan Evans</i>	
<b>24</b>	<b>Risk Communication in Health</b>	<b>621</b>
	<i>Nicolai Bodemer · Wolfgang Gaissmaier</i>	
<b>25</b>	<b>Risk Perception and Societal Response</b>	<b>661</b>
	<i>Lennart Sjöberg</i>	
<b>26</b>	<b>The Role of Feelings in Perceived Risk</b>	<b>677</b>
	<i>Melissa L. Finucane</i>	
<b>27</b>	<b>Emotion, Warnings, and the Ethics of Risk Communication</b>	<b>693</b>
	<i>Ross Buck · Rebecca Ferrer</i>	
<b>28</b>	<b>Cultural Cognition as a Conception of the Cultural Theory of Risk</b>	<b>725</b>
	<i>Dan M. Kahan</i>	
<b>29</b>	<b>Tools for Risk Communication</b>	<b>761</b>
	<i>Britt-Marie Drottz-Sjöberg</i>	

### Part 5 Risk Ethics

<b>30</b>	<b>Ethics and Risk</b>	<b>791</b>
	<i>Douglas MacLean</i>	
<b>31</b>	<b>Toward a Premarket Approach to Risk Assessment to Protect Children</b>	<b>805</b>
	<i>Carl F. Cranor</i>	
<b>32</b>	<b>Moral Emotions as Guide to Acceptable Risk</b>	<b>819</b>
	<i>Sabine Roeser</i>	
<b>33</b>	<b>Risk and Virtue Ethics</b>	<b>833</b>
	<i>Allison Ross · Nafsika Athanassoulis</i>	
<b>34</b>	<b>Risk and Trust</b>	<b>857</b>
	<i>Philip J. Nickel · Krist Vaesen</i>	
<b>35</b>	<b>Risk and Responsibility</b>	<b>877</b>
	<i>Ibo van de Poel · Jessica Nihlén Fahlquist</i>	

<b>36</b>	<b>What Is a Fair Distribution of Risk? . . . . .</b>	<b>909</b>
	<i>Madeleine Hayenjelm</i>	
<b>37</b>	<b>Intergenerational Risks . . . . .</b>	<b>931</b>
	<i>Lauren Hartzell-Nichols</i>	
<b>38</b>	<b>The Precautionary Principle . . . . .</b>	<b>961</b>
	<i>Marko Ahteesuu · Per Sandin</i>	
<b>39</b>	<b>The Capability Approach in Risk Analysis . . . . .</b>	<b>979</b>
	<i>Colleen Murphy · Paolo Gardoni</i>	

## **Part 6 Risk in Society**

<b>40</b>	<b>Sociology of Risk . . . . .</b>	<b>1001</b>
	<i>Rolf Lidskog · Göran Sundqvist</i>	
<b>41</b>	<b>Risk and Gender: Daredevils and Eco-Angels . . . . .</b>	<b>1029</b>
	<i>Misse Wester</i>	
<b>42</b>	<b>Risk and Soft Impacts . . . . .</b>	<b>1049</b>
	<i>Tsjalling Swierstra · Hedwig te Molder</i>	
<b>43</b>	<b>Risk and Technology Assessment . . . . .</b>	<b>1067</b>
	<i>Rinie van Est · Bart Walhout · Frans Brom</i>	
<b>44</b>	<b>Risk Governance . . . . .</b>	<b>1093</b>
	<i>Marijke A. Hermans · Tessa Fox · Marjolein B. A. van Asselt</i>	
<b>45</b>	<b>EU Risk Regulation and the Uncertainty Challenge . . . . .</b>	<b>1119</b>
	<i>Marjolein B. A. van Asselt · Ellen Vos</i>	
<b>46</b>	<b>Risk Management in Technocracy . . . . .</b>	<b>1137</b>
	<i>Val Dusek</i>	
<b>Index</b>	<b>1165</b>	

# Editors

## **Sabine Roeser (Editor-in-Chief)**

Philosophy Department  
Faculty of Technology,  
Policy and Management  
Delft University of Technology  
Delft  
The Netherlands  
and  
Philosophy Department  
University of Twente  
Enschede  
The Netherlands

## **Rafaela Hillerbrand**

Human Technology Center  
RWTH Aachen University  
Aachen  
Germany

## **Per Sandin**

Department of Plant Physiology and Forest  
Genetics  
Swedish University of Agricultural Sciences  
Uppsala  
Sweden

## **Martin Peterson**

Section for Philosophy and Ethics  
Eindhoven University of Technology  
Eindhoven  
The Netherlands



# List of Contributors

## **John Adams**

Department of Geography  
University College London  
London  
UK

## **Marko Ahteesuu**

Public Choice Research Centre (PCRC)  
University of Turku  
Turku  
Finland

## **Nafsika Athanassoulis**

Oswestry  
UK

## **Claudia Basta**

Section Philosophy  
Delft University of Technology  
3TU. Centre for Ethics and Technology  
Delft  
The Netherlands  
and  
Land Use Planning Group  
Wageningen University and Research  
Centre  
Wageningen  
The Netherlands

## **Claus Beisbart**

Fakultät Humanwissenschaften und  
Theologie  
Institut für Philosophie und  
Politikwissenschaft  
Technische Universität Dortmund  
Dortmund  
Germany

## **Nicolai Bodemer**

Harding Center for Risk Literacy  
Max Planck Institute for Human  
Development  
Berlin  
Germany

## **Frans Brom**

Department of Technology Assessment  
Rathenau Institute  
The Hague  
The Netherlands

## **Ross Buck**

Department of Communication Sciences  
University of Connecticut  
Storrs, CT  
USA

## **Carl F. Cranor**

Department of Philosophy  
University of California  
Riverside, CA  
USA

## **Amy Donovan**

Department of Geography  
University of Cambridge  
Cambridge  
UK

## **Britt-Marie Drott-Sjöberg**

Risk Psychology, Environment and Safety  
Research Group  
Department of Psychology  
Norwegian University of Science and  
Technology (NTNU)  
Trondheim  
Norway

**Val Dusek**

Department of Philosophy  
University of New Hampshire  
Durham, NH  
USA

**Louis Eeckhoudt**

IÉSEG School of Management  
Lille  
France  
and  
The Center for Operations Research and  
Econometrics (CORE)  
Université catholique de Louvain  
Louvain  
Belgium

**Dylan Evans**

School of Medicine  
University College Cork  
Cork  
Ireland

**Rebecca Ferrer**

Behavioral Research Program  
Division of Cancer Control and Population  
Sciences  
National Cancer Institute  
Rockville, MD  
USA

**Melissa L. Finucane**

East-West Center  
Honolulu, HI  
USA

**Tessa Fox**

Department of Technology and Society  
Studies  
Faculty of Arts & Social Sciences  
Maastricht University  
Maastricht  
The Netherlands

**Wolfgang Gaissmaier**

Harding Center for Risk Literacy  
Max Planck Institute for Human  
Development  
Berlin  
Germany

**Paolo Gardoni**

Zachry Department of Civil Engineering  
Texas A&M University  
College Station, TX  
USA

**Till Grüne-Yanoff**

Helsinki Collegium of Advanced Studies  
University of Helsinki  
Helsinki  
Finland

**Sven Ove Hansson**

Division of Philosophy  
Royal Institute of Technology  
Stockholm  
Sweden

**Lauren Hartzell-Nichols**

Program on Values in Society  
University of Washington  
Seattle, WA  
USA

**Madeleine Hayenjelm**

Department of Philosophy  
University College London  
London  
UK

**Marijke A. Hermans**

Department of Technology and Society  
Studies  
Faculty of Arts & Social Sciences  
Maastricht University  
Maastricht  
The Netherlands

<b>Rafaela Hillerbrand</b> Human Technology Center & Institute for Philosophy RWTH Aachen University Aachen Germany	<b>Niklas Möller</b> University of Cambridge Cambridge UK
<b>Karsten Klint Jensen</b> Danish Centre for Bioethics and Risk Assessment Institute of Food and Resource Economics University of Copenhagen Frederiksberg C Denmark	<b>Colleen Murphy</b> Department of Philosophy Texas A&M University College Station, TX USA
<b>Dan M. Kahan</b> Yale Law School Yale University New Haven, CT USA	<b>Philip J. Nickel</b> Department of Philosophy and Ethics School of Innovation Sciences Eindhoven University of Technology Eindhoven The Netherlands
<b>Rolf Lidskog</b> Centre for Urban and Regional Studies Örebro University Örebro Sweden	<b>Jessica Nihlén Fahlquist</b> Delft University of Technology Delft The Netherlands and Royal Institute of Technology Stockholm Sweden
<b>Henri Loubergé</b> Department of Economics Geneva Finance Research Institute (GFRI) University of Geneva Swiss Finance Institute Geneva Switzerland	<b>Martin Peterson</b> Section for Philosophy and Ethics Eindhoven University of Technology Eindhoven The Netherlands
<b>Douglas MacLean</b> Department of Philosophy University of North Carolina Chapel Hill, NC USA	<b>Annette Rid</b> Institute of Biomedical Ethics University of Zurich Zurich Switzerland
<b>Erika Mansnerus</b> Health and Social Care London School of Economics London UK	<b>Hauke Riesch</b> Judge Business School University of Cambridge Cambridge UK

**Sabine Roeser**

Philosophy Department  
Delft University of Technology  
Delft  
The Netherlands  
and  
Philosophy Department  
University of Twente  
Enschede  
The Netherlands

**Allison Ross**

London  
UK

**Nils-Eric Sahlin**

Department of Medical Ethics  
Lund University  
Lund  
Sweden  
and  
Center for Philosophy of Science  
University of Pittsburgh  
Pittsburgh, PA  
USA

**Per Sandin**

Department of Plant Physiology and Forest  
Genetics  
Swedish University of Agricultural Sciences  
Uppsala  
Sweden

**Lennart Sjöberg**

Center for Risk Research  
Center for Media and Economic Psychology  
Marketing and Strategy Department  
Stockholm School of Economics  
Stockholm  
Sweden  
and  
Center for Risk Psychology, Environment,  
and Safety  
Department of Psychology  
Norwegian University of Science and  
Technology  
Trondheim  
Norway

**Marjan Sjerps**

Department of Science, Interdisciplinary  
Investigations, Statistics, and Knowledge  
Management  
Netherlands Forensic Institute  
The Hague  
The Netherlands

**Reinoud D. Stoel**

Department of Science, Interdisciplinary  
Investigations, Statistics, and Knowledge  
Management  
Netherlands Forensic Institute  
The Hague  
The Netherlands

**Göran Sundqvist**

Centre for Technology, Innovation and  
Culture  
University of Oslo  
Oslo  
Norway

**Tsjalling Swierstra**

Department of Philosophy  
University of Maastricht  
Maastricht  
The Netherlands

**Behnam Taebi**

Department of Philosophy  
Delft University of Technology  
Delft  
The Netherlands

**Peter R. Taylor**

Oxford Martin School  
University of Oxford  
Old Indian Institute  
Oxford  
UK

<b>Hedwig te Molder</b> Faculty of Behavioral Sciences, Science Communication University of Twente/Wageningen University Enschede/Wageningen The Netherlands	<b>Ferdinand M. Vieider</b> Institut für Volkswirtschaftslehre Ludwig-Maximilians-Universität München Munich Germany
<b>Stefan T. Trautmann</b> Tilburg Institute of Behavioral Economics Research Department of Social Psychology & Center Tilburg University Tilburg The Netherlands	<b>Ellen Vos</b> Faculty of Law Maastricht University Maastricht The Netherlands
<b>Krist Vaesen</b> Department of Philosophy and Ethics Eindhoven University of Technology Eindhoven The Netherlands	<b>Bart Walhout</b> Department of Technology Assessment Rathenau Institute The Hague The Netherlands
<b>Marjolein B. A. van Asselt</b> Department of Technology and Society Studies Faculty of Arts & Social Sciences Maastricht University Maastricht The Netherlands	<b>John Weckert</b> Centre for Applied Philosophy and Public Ethics (CAPPE) Charles Sturt University Canberra, ACT Australia
<b>Ibo van de Poel</b> Delft University of Technology Delft The Netherlands	<b>Paul Weirich</b> Department of Philosophy University of Missouri Columbia, MO USA
<b>Rinie van Est</b> Department of Technology Assessment Rathenau Institute The Hague The Netherlands	<b>John R. Welch</b> Department of Philosophy Saint Louis University, Madrid Campus Madrid Spain
	<b>Misse Wester</b> Division of Philosophy The Royal Institute of Technology Stockholm Sweden



# 1 Introduction to Risk Theory

Sabine Roeser<sup>1,2</sup> · Rafaela Hillerbrand<sup>3</sup> · Per Sandin<sup>4</sup> · Martin Peterson<sup>5</sup>

<sup>1</sup>Delft University of Technology, Delft, The Netherlands

<sup>2</sup>University of Twente, Enschede, The Netherlands

<sup>3</sup>RWTH Aachen University, Aachen, Germany

<sup>4</sup>Swedish University of Agricultural Sciences, Uppsala, Sweden

<sup>5</sup>Eindhoven University of Technology, Eindhoven, The Netherlands

<b>Introduction</b> .....	3
<b>Part 1: General Issues in Risk Theory</b> .....	3
Sven Ove Hansson: A Panorama of the Philosophy of Risk .....	3
Niklas Möller: The Concepts of Risk and Safety .....	4
Hauke Riesch: Levels of Uncertainty .....	4
<b>Part 2: Specific Risks</b> .....	5
Louis Eeckhoudt and Henri Loubergé: The Economics of Risk: A (Partial) Survey .....	5
Reinoud D. Stoel and Marjan Sjerps: Interpretation of Forensic Evidence .....	5
John Weckert: Risks and Scientific Responsibilities in Nanotechnology .....	5
Annette Rid: Risk and Risk-Benefit Evaluations in Biomedical Research .....	6
Erika Mansnerus: Understanding and Governing Public Health Risks by Modeling .....	6
John Adams: Management of the Risks of Transport .....	6
Claudia Basta: Risk and Spatial Planning .....	7
Behnam Taebi: Intergenerational Risks of Nuclear Energy .....	7
Rafaela Hillerbrand: Climate Change as Risk? .....	8
Amy Donovan: Earthquakes and Volcanoes: Risk from Geophysical Hazards .....	8
<b>Part 3: Decision Theory and Risk</b> .....	8
Claus Beisbart: A Rational Approach to Risk? Bayesian Decision Theory .....	10
Karsten Klint Jensen: A Philosophical Assessment of Decision Theory .....	10
Peter R. Taylor: The Mismeasure of Risk .....	10
Nils-Eric Sahlin: Unreliable Probabilities, Paradoxes, and Epistemic Risks .....	11
Till Grüne-Yanoff: Paradoxes of Rational Choice Theory .....	11
Paul Weirich: Multi-Attribute Approaches to Risk .....	11
John R. Welch: Real-Life Decisions and Decision Theory .....	12
Stefan T. Trautmann and Ferdinand M. Vieider: Social Influences on Risk Attitudes: Applications in Economics .....	12
<b>Part 4: Risk Perception</b> .....	12
Dylan Evans: Risk Intelligence .....	13
Nicolai Bodemer and Wolfgang Gaissmaier: Risk Communication in Health .....	13
Lennart Sjöberg: Risk Perception and Societal Response .....	13

Melissa L. Finucane: The Role of Feelings in Perceived Risk .....	14
Ross Buck and Rebecca Ferrer: Emotion, Warnings, and the Ethics of Risk Communication .....	14
Dan M. Kahan: Cultural Cognition as a Conception of the Cultural Theory of Risk .....	15
Britt-Marie Drott-Sjöberg: Tools for Risk Communication .....	15
 <b>Part 5: Risk Ethics .....</b>	 <b>16</b>
Douglas MacLean: Ethics and Risk .....	16
Carl F. Cranor: Toward a Premarket Approach to Risk Assessment to Protect Children .....	16
Sabine Roeser: Moral Emotions as Guide to Acceptable Risk .....	17
Allison Ross and Nafsika Athanassoulis: Risk and Virtue Ethics .....	17
Philip J. Nickel and Krist Vaesen: Risk and Trust .....	18
Ibo van de Poel and Jessica Nilén Fahlquist: Risk and Responsibility .....	18
Madeleine Hayenjelm: What Is a Fair Distribution of Risk? .....	18
Lauren Hartzell-Nichols: Intergenerational Risks .....	19
Marko Ahteesuu and Per Sandin: The Precautionary Principle .....	19
Colleen Murphy and Paolo Gardoni: The Capability Approach in Risk Analysis .....	19
 <b>Part 6: Risk in Society .....</b>	 <b>20</b>
Rolf Lidskog and Göran Sundqvist: Sociology of Risk .....	20
Misse Wester: Risk and Gender: Daredevils and Eco-Angels .....	21
Tsjalling Swierstra and Hedwig te Molder: Risk and Soft Impacts .....	21
Rinie van Est, Bart Walhout, and Frans Brom: Risk and Technology Assessment .....	21
Marijke A. Hermans, Tessa Fox, and Marjolein B. A. van Asselt: Risk Governance .....	22
Marjolein B. A. van Asselt and Ellen Vos: EU Risk Regulation and the Uncertainty Challenge .....	22
Val Dusek: Risk Management in Technocracy .....	23
 <b>Conclusion .....</b>	 <b>23</b>

## Introduction

---

Risk is an important topic in contemporary society. People are confronted with risks from financial markets, nuclear power plants, natural disasters and privacy leaks in ICT systems, to mention just a few of a sheer endless list of areas in which uncertainty and risk of harm play an important role. It is in that sense not surprising that risk is studied in fields as diverse as mathematics and natural sciences but also psychology, economics, sociology, cultural studies, and philosophy. The topic of risk gives rise to concrete problems that require empirical investigations, but these empirical investigations need to be structured by theoretical frameworks. This handbook offers an overview of different approaches to risk theory, ranging from general issues in risk theory to risk in practice, from mathematical approaches in decision theory to empirical research of risk perception, to theories of risk ethics and to frameworks on how to arrange society in order to deal appropriately with risk.

Risk theory provides frameworks that can contribute to mitigating risks, coming to grips with uncertainty, and offering ways to organize society in such a way that the unexpected and unknown can be anticipated or at least dealt with in a reasonable and ethically acceptable way. This handbook reflects the current state of the art in risk theory, by bringing together scholars from various disciplines who review the topic of risk from different angles.

## Part 1: General Issues in Risk Theory

---

Theoretical reflection about risk gives rise to various general issues. What is the relation between risk research and philosophy? What can the two disciplines learn from each other? How are the concepts of risk and safety interrelated? Which different levels of uncertainty can be distinguished? Part one of this handbook discusses these general issues in risk theory.

### Sven Ove Hansson: A Panorama of the Philosophy of Risk

---

Sven Ove Hansson's chapter provides for an overview of the contributions that different philosophical subdisciplines and risk theory can provide for each other. He starts out with discussing the potential contributions of philosophy to risk theory. These contributions concern terminological clarification, argumentation theory, and the fact–value distinction. It is philosophy's “core business” to provide for terminological clarification. In the area of risk theory, philosophical theories can shed light on the multifaceted concepts of risk and safety. Philosophical argumentation theory can draw attention to common fallacies in reasoning about risk. In the philosophical tradition, there are intricate debates about the relationship between facts and values that can contribute to more careful and nuanced discussions about facts and values in risk analysis, for example, by making implicit value judgments explicit. These are areas in which existing philosophical theories can be applied rather straightforwardly. However, there are other issues in risk theory that give rise to new philosophical problems and require new philosophical approaches in virtually all areas of philosophy. This is due to the fact that most traditional philosophical approaches are based on deterministic assumptions. Thinking about risk and uncertainty requires radically different philosophical

theories. Hence, the topic of risk can lead to new philosophical theories that are at the same time directly practically relevant, as our real-life world is one that is characterized by risk and uncertainty.

## Niklas Möller: The Concepts of Risk and Safety

---

Niklas Möller provides for an analysis of the concepts of risk and safety. Möller distinguishes three major approaches in empirically oriented risk theory: the scientific, the psychological, and the cultural approach, respectively. Philosophical approaches connect with each of these approaches in specific ways. Möller then distinguishes between at least five common usages of the notion “risk” and shows how each usage can be connected to a different approach in risk research. He emphasizes the importance of distinguishing between factual and normative uses of the notion of risk. Möller continues with the question whether safety is the antonym of risk. He argues that this is not necessarily the case; for example, these words can have different connotations. Möller then discusses various ethical aspects of risk that are elaborated on in more detail in Part 5 of this handbook, on Risk Ethics. Möller’s unique contribution to risk ethics is to argue that risk is a “thick concept”; that is, a concept that does not only have descriptive aspects that are the subject of scientific investigations, but that also has normative or evaluative aspects, which require ethical reflection. Möller discusses and rejects, on philosophical grounds, various claims by social scientists to the socially constructed nature of risk that is supposed to follow from its inherently normative nature.

## Hauke Riesch: Levels of Uncertainty

---

Risk science seems to be a paradigm of interdisciplinary research: Risk unites disciplines. But every discipline seems to denote something else with the umbrella term risk. This dilemma is the starting point of Hauke Riesch’s contribution on the “Levels of Uncertainty.” Hauke Riesch analyzes various uses of the terms risk and uncertainty. He attributes differences not so much to imperfect or sloppy use of the terms. Rather, he argues that these differences are a symptom of the fact that different scientists are interested in different aspects of risk. Therefore, there is not much point in criticizing someone for using vague or different notions of risk. Riesch conceptualizes risk as uncertainty of an event happening whose outcome may be severe. Riesch argues that concerning the uncertainty aspect of risk we can distinguish the following six questions which are not mutually independent: Why are we uncertain? Who is uncertain? How is uncertainty represented? How do people react to uncertainty? How do we understand uncertainty? What exactly are we uncertain about? Within this multidimensional map, Riesch divides the objects of uncertainty into five layers: uncertainty of the outcome, uncertainty about the parameters as well as uncertainty about the model itself, uncertainty about acknowledged inadequacies and implicitly made assumptions, and uncertainty about the unknown inadequacies. These layers relate to different concerns of different disciplines. The expert discourse commonly focuses on one of these levels, but this distorts the way people perceive particular risks, because higher level uncertainties still exist. Riesch illustrates through various case studies – lottery, bad eating habits, CCS (carbon capture storage), and climate

change – how all five levels of uncertainty are always present, but differently important. Riesch's multidimensional classification provides some useful information for risk communication in order to convey information on other levels of uncertainty that people find important.

## Part 2: Specific Risks

---

What brings about reflections on risk is the necessity to react to real natural or anthropogenic hazards. The second part of this book addresses natural, technological, and societal risks from the perspective of the natural and social sciences by also incorporating insights from the humanities and mathematical sciences.

### Louis Eeckhoudt and Henri Loubergé: The Economics of Risk: A (Partial) Survey

---

Louis Eeckhoudt and Henri Loubergé give a historical overview of how risk thinking has developed in the mainstream model of economics, that is, that of expected utility theory. The authors argue that despite Daniel Bernoulli's foundational work on this model in the eighteenth century, those ideas were not formulated in precise terms until von Neumann and Morgenstern's Expected Utility Theorem in the late 1940s. These ideas were further developed by Friedman and Savage, Arrow, and Pratt. Eeckhoudt and Loubergé introduce the concept of general equilibrium and discuss issues of risk distribution among individuals. The authors end their discussion with brief illustrations of applications of the theory from the fields of finance and insurance.

### Reinoud D. Stoel and Marjan Sjerps: Interpretation of Forensic Evidence

---

Reinoud D. Stoel and Marjan Sjerps write about the interpretation of forensic evidence. Since absolute certainty regarding the guilt of a suspect is unattainable, the question has to be put in terms of probability. They go on to describe how this issue can be approached using the Likelihood Ratio and how this applies to both forensic experts and legal decision makers. They propose further research focusing on methods for computing quantitative Likelihood Ratios for different forensic disciplines and also how different pieces of evidence combine, for example, using Bayesian networks.

### John Weckert: Risks and Scientific Responsibilities in Nanotechnology

---

John Weckert writes about the risks of nanotechnology, and using this field as an example he discusses scientists' responsibility. He presents four generations of nanotechnology and the risks associated with each generation. In particular, he emphasizes five risks that have been discussed in relation to nanotechnology: health/environmental risks related to nanoparticles, "grey goo," threats to privacy, cyborgs, and the possibility of a "nanodivide" between the

developed and the developing world. He goes on to delineate two models of science, the linear and the social mode, where the latter focuses on the inevitable value-ladenness of science. He also discusses four different interfaces between science and ethics. As regards the responsibility of scientists involved in nano research, he argues that these responsibilities differ between the different risks, and that on the linear model scientists are not free from responsibility. Weckert calls for more interdisciplinary research in the future, based on sound science and knowledge of actual technological developments.

### **Annette Rid: Risk and Risk-Benefit Evaluations in Biomedical Research**

---

Annette Rid critically reviews recent debates about risk-benefit evaluations in biomedical research. To determine whether potential new interventions in clinical care, new drugs, or basic science research in biomedicine have a net benefit, risks that participants take have to be balanced against the positive effects for potential patients and society as a whole. Rid presents and evaluates the four existing ethical frameworks for risk-benefit evaluations in medical ethics: component analysis, the integrative approach, the agreement principle, and the net risks test. She contends that net risks tests are superior to alternative approaches, but fail to offer guidance for evaluating the ethical acceptability of risks that participants are exposed to for research purposes only. This leaves two of the fundamental problems of risk-benefit evaluations in research largely unaddressed: How to weigh the risks to the individual research participant against the potential social value of the knowledge to be gained from a study, and how to set upper limits of acceptable research risk. Discussions about the “minimal” risk threshold in research with participants who cannot consent go some way to specifying upper risk limits in this context. However, these discussions apply only to a small portion of research studies.

### **Erika Mansnerus: Understanding and Governing Public Health Risks by Modeling**

---

Erika Mansnerus discusses how risk is perceived and expressed through the computational models that are increasingly used to govern and understand public health risk. She discusses a case study on infectious disease epidemiology to illustrate how models are used for explanation-based and scenario-building predictions in order to anticipate the risks of infections. Mansnerus analyzes the tension that arises when model-based estimates exemplify the population-level reasoning of public health risks, but have restricted capacity to address risks on an individual level. Mansnerus provides an alternative account in order to overcome the limitations of computational tools in the governance of public health risks. This richer picture on risk goes beyond the pure probabilistic realm of mathematical risk modeling.

### **John Adams: Management of the Risks of Transport**

---

John Adams discusses whether transport safety regulations indeed reduce risks of, for example, fatal car accidents. With the exception of the seat belt law, major reductions in road fatalities

over the past decades cannot be linked to changes in transport safety laws in a straightforward way. So what is the use of these laws and why do they fail? Adams introduces the reader to two models of how risk managers may deal with risks, a cost–benefit model and a model underlying risk compensation strategies. Both models focus on different risk. The former focuses on what Adams terms “risks perceived through science” and models road users as ignorant, obedient objects. In contrast, the alternative model pictures road users as vigilant and responsive subjects and thus can also take into account directly perceived risks and what Adams terms “virtual risks.” Adams applies Mary Douglas’ cultural theory of risk to road safety issues in order to provide an explanation of why different societies perceive risks very differently. Adams shows how this leads to two very different strategies to manage and reduce the risks of transport in the two countries with the best road safety records worldwide, the Netherlands and Sweden.

## Claudia Basta: Risk and Spatial Planning

---

Claudia Basta challenges the way societies currently deal with site-specific hazardous technologies. The location of hazardous facilities is rarely an uncontroversial issue. Basta describes how this issue is dealt with in selected European countries and explains their differences as a reflection of nonexplicit, but retraceable, underlying ethical theories. Basta suggests a possible synergy between spatial planning practices and ethical theories by proposing a theoretical framework which may guide spatial planning processes before and beyond unavoidable contextual features. In particular, following the Rawlsian theory of justice (Rawls 1971) as transposed to spatial planning theories by Moroni, Basta proposes an understanding of “spatial safety” as a primary spatial good. Spatial planning practice is thus perceived primarily as a practice of distributing the maximum possible amount of the primary good of spatial safety in society equally up to the lowest societal level. She contends that approaching any siting controversy as a NIMBY (not-in-my-backyard) case is not only incorrect, but also dangerously instrumental. The case study of a planned CCS facility at Barendrecht illustrates this point.

## Behnam Taebi: Intergenerational Risks of Nuclear Energy

---

Behnam Taebi discusses the intergenerational risk of nuclear waste disposal. The accident in the Fukushima II nuclear power plant in Japan in 2011 gives rise to a renewed discussion about the civic use of nuclear power. In addition to the risk of a nuclear meltdown, nuclear power plants put seemingly unduly high burdens on future generations: The longevity of the toxic and radioactive waste seems to require geological disposals that are faced with immeasurable uncertainties concerning the stability of rock formation over large timescales (a few thousand years). The relatively new technological possibility of Partition and Transmutation (PT) may provide an alternative to that and put long-term surface storage in a new light. Taebi not only examines this new technology as an alternative to geological disposal, but also addresses central aspects of intergenerational ethics, namely the principles of intergenerational equity. Taebi scrutinizes the notion of diminishing responsibility over time, which is an important notion in

nuclear waste policies, and rejects the moral legitimacy of distinguishing between future people. In this sense Taebi not only provides a thorough account on how to deal with long-life radioactive waste, but also reflects on our present obligations toward future generations.

### Rafaela Hillerbrand: Climate Change as Risk?

---

Rafaela Hillerbrand analyzes the types of uncertainties involved in climate modeling and discusses whether common decision approaches based on the precautionary principle or on the maximization of expected utility are capable of incorporating the uncertainty inevitably involved in climate modeling. The author contends that in the case of decision making about climate change, unquantified uncertainties can neither be ignored, nor can they be reduced to quantified uncertainties, by assigning subjective probabilities. These insights reveal central problems, as they imply that the commonly used elementary as well as probabilistic decision approaches are not applicable in this case. The chapter argues that the epistemic problems involved in modeling the climate system are generic for modeling complex systems. Possible parts of future research to circumvent these problems are adumbrated.

### Amy Donovan: Earthquakes and Volcanoes: Risk from Geophysical Hazards

---

The recent earthquake in Japan revealed the vulnerability of a high-tech society to natural hazards. Amy Donovan calls for a genuinely interdisciplinary study of volcano and seismic risk if we want to be better forewarned against such risks. A wide range of disciplines, spanning both social and physical sciences, is involved in research into geophysical hazards, in order to predict, prepare for, and communicate about events like the one in Japan. However, only few holistic approaches to geophysical risks exist. Most information comes from the natural sciences that tend to focus on hazard assessment, while the social sciences focus on vulnerability reduction and risk communication. The contribution of Amy Donovan fills this gap. Donovan examines the social context both of the scientific research of the natural hazards, and of the hazards themselves, proposing a holistic and context-based approach to understanding risk. In the context of seismic risk, the uncertainty of the scientific methods coupled with the procedures involved in mitigating these risks and the need to involve populations in the preparation, leads to a snowballing of uncertainty and indeterminacy from the scientific domain through the policy domain and into the wider public.

## Part 3: Decision Theory and Risk

---

As illustrated by the interdisciplinary nature of the contributions to this handbook, risk theory is a very broad field of research. Eight chapters explore the links between risk theory and *decision theory*.

Very briefly put, decision theory is the theory of rational decision making. A key assumption accepted by nearly all decision theorists is that it is essential to distinguish between

*descriptive* claims about how people actually make risky decisions, and *normative* claims about how it would be rational to take such decisions. In recent years, the term decision theory has primarily been used for referring to the study of normative claims about rational decision making. In what follows, we shall follow this convention and reserve the term “decision theory” for the study of normative hypotheses about what rationality demands of us.

The notion of rationality researched in decision theory is best described as means–ends rationality. The rational decision maker has certain beliefs about what the effects of various risky options could be, and also has a set of desires about what he or she wants to achieve. The key question is then how we should best combine these beliefs and desires into a decision.

This so-called belief–desire model of rationality goes back at least to the Scottish eighteenth-century philosopher David Hume. According to Hume, the best explanation of why people behave as they do – and are willing to accept the risks they do accept – is that we all have certain beliefs and desires that determine our actions (cf. Hume 1740/1967). In the belief–desire model, a basic criterion of a rational decision is that it must accurately reflect the agent’s beliefs and desires, no matter what these beliefs and desires happen to be about. But how should such a theory of rationally permissible structures of beliefs and desires be spelled out in detail?

The mainstream view among contemporary decision theorists is that rational agents are allowed to let whatever beliefs and desires they so wish guide their decisions, as long as those beliefs and desires are compatible with the principle of maximizing expected utility. The principle of maximizing expected utility takes the total value of an act to equal the sum total of the values of its possible outcomes weighted by the probability for each outcome. The values assigned to an outcome are determined by the decision maker’s desires, whereas the probabilities are determined by his or her beliefs about how likely the outcomes are to materialize.

An important step in the development of modern decision theory was the development of axiomatic accounts of the principle of maximizing expected utility. The first such axiomatization was sketched by Frank Ramsey in his paper *Truth and Probability*, written in 1926 but published posthumously in 1931. Ramsey (1931) formulated eight axioms for how rational decision makers should form preferences among uncertain prospects. One of the many points in his paper is that a decision maker who behaves in accordance with the eight axioms will act in a way that is *compatible* with the principle of maximizing expected value, in the sense that he or she will implicitly assign numerical probabilities and values to outcomes. Note that it does not follow from this that the decision maker’s choices were *actually triggered* by these implicit probabilities and utilities. This indirect approach to rational decision making is extremely influential in the contemporary literature. In 1954, Leonard Savage put forward roughly the same idea in his influential book *The Foundations of Statistics* (Savage 1954).

An equally important book in the recent history of decision theory is John von Neumann and Oskar Morgenstern’s book *Theory of Games and Economic Behavior* (1944). They used the notion of a “lottery” for developing a linear measure of an outcome’s utility. In von Neumann and Morgenstern’s vocabulary, a lottery is a probabilistic mixture of outcomes. For instance, an entity such as “a fifty-fifty chance of winning either \$1,000 or a trip to Miami” is a lottery. The upshot of their utility theory is that every decision maker whose preferences over a very large set of lotteries conform to a small set of axioms implicitly assigns numerical utilities to outcomes, and also implicitly acts in accordance with the principle of maximizing expected utility.

The chapters in this section discuss various aspects of, ideas behind, and problems with decision theory in the context of risk.

## Claus Beisbart: A Rational Approach to Risk? Bayesian Decision Theory

---

Claus Beisbart's contribution clarifies the concept of utility maximization as the core notion of rationality in Bayesian decision theory. For Bayesians, a rational approach to risk depends on the agents' utilities and subjective probabilities that measure the strength of their desires and beliefs, respectively. Two questions are commonly put forward by critics of Bayesian decision theory: Why do or why should rational agents maximize expected utility? And how can the strength of an agent's desire and belief be measured? The classical answers by Bayesians to these questions are commonly put in the form of representation theorems. These theorems show that – under certain assumptions – an agent's beliefs and desires can indeed be represented in terms of numerical probabilities and utilities. Beisbart presents several different ways to obtain a presentation theorem: the classical approach of von Neumann-Morgenstern, its modification by Anscombe and Aumann, the approach of Ramsey, Bolker-Jeffrey theory, and Savage's account, the latter possibly being closest to how risk is actually dealt with in technology assessment. The chapter concludes with a discussion of major controversies concerning Bayesian decision theory.

## Karsten Klint Jensen: A Philosophical Assessment of Decision Theory

---

Karsten Klint Jensen starts his chapter with drawing a distinction between classical decision theory and modern axiomatic decision theory. He then goes on to give an overview of Savage's axiomatization of the principle of maximizing expected utility. This all leads up to a discussion of what *sort of problem* decision theory really aims at solving. Several possible answers are considered. Jensen points out that the perhaps most plausible interpretation of what modern decision theories are up to is to try to develop theories of what counts as personal good and how such personal goods can be aggregated intrapersonally.

## Peter R. Taylor: The Mismeasure of Risk

---

Peter R. Taylor challenges the classical approaches in risk management and risk assessment, thereby mainly, but not exclusively, focusing on how the insurance industry deals with risk. Taylor argues that already our embosomed definition of risk, which pictures risk as a quantitative measure combining, in one way or the other, harm and likelihood of the hazardous event, falls short of describing realistic events like the recent disasters that caught world headlines – tsunamis, volcanic ash clouds, or financial crashes. It is argued that simple measures of risk that, like for example in the Bayesian approach, focus merely on the mean expected harm, may be poor guides for dealing with real-world risk. Taylor outlines how a more complex risk assessment may incorporate what Rumsfeld termed the unknown unknowns or Taleb the black swans: that is, events that are not considered in the probability space. Taylor shows how model risk, that is, the risk that the underlying model of the, say, physical hazard is simply inadequate, may be quantified and how multiple measures and thresholds may be implemented. The author further examines if we could also tackle what he calls ROB, "risk outside the box," that is, risks that are not considered by the underlying risk model.

## Nils-Eric Sahlin: Unreliable Probabilities, Paradoxes, and Epistemic Risks

Nils-Eric Sahlin's chapter focuses on the quality of the information on which a decision is based. By definition, if the information at hand when a decision is taken is unreliable, then the epistemic uncertainty is high. Suppose, for instance, that for one reason or another tomorrow's weather is important for a decision that you are about to take. Now consider the scenario in which you have received a detailed weather forecast from a professional meteorologist according to which the probability for rain tomorrow is 50% and compare this with a scenario in which your 3-year-old son, who knows nothing about meteorology, tells you that he thinks the probability for rain tomorrow is 50%. Intuitively, you as a decision maker seem to be better off from an epistemic point of view in the first scenario compared to the second, but exactly how should this difference be accounted for from a decision theoretical point of view? How do we take intuitions about epistemic risks into account in our theories of rational decision making? This is the key question that drives Sahlin's chapter.

## Till Grüne-Yanoff: Paradoxes of Rational Choice Theory

Till Grüne-Yanoff's chapter gives an overview of some of the many well-known paradoxes that have played an important role in the development of decision theory. The very first such paradox was the St Petersburg paradox, formulated by the Swiss mathematician Daniel Bernoulli (1700–1782), who worked in St Petersburg for a couple of years at the beginning of the eighteenth century. The St Petersburg paradox, which is still being discussed by decision theorists, is derived from a game known as the St Petersburg game: A fair coin is tossed until it lands heads up. The player then wins  $2^n$  dollars, where  $n$  is the number of times the coin was tossed. Hence, if the coin lands heads up in the first toss, the player wins 2 dollars, but if it lands heads up on, say, the fourth toss, the player wins  $2^4 = 2 \cdot 2 \cdot 2 \cdot 2 = 16$  dollars. How much should a rational person be willing to pay for getting the opportunity to play this game? Clearly, the expected monetary payoff is infinite, because  $\frac{1}{2} \cdot 2 + \frac{1}{4} \cdot 4 + \frac{1}{8} \cdot 8 + \dots = 1 + 1 + 1 + \dots = \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n \cdot 2^n = \infty$ . However, to pay, say, a million dollars for playing a game in which one is very likely to win just 2, 4, or 8 dollars seems absurd. As Grüne-Yanoff points out, it is not as easy to resolve this and other paradoxes as many people have thought.

## Paul Weirich: Multi-Attribute Approaches to Risk

Paul Weirich considers the special type of situation that arises if many different features of a decision are considered to be relevant by the decision maker. In a single-attribute approach to decision theory, all possible outcomes are compared on one and the same scale. Imagine, for instance, that you are about to buy a new car. Some cars are more expensive than others, but they are also safer. How many dollars is it worth to pay for extra safety? The multi-attribute approach attempts to avoid the idea that money and human welfare are somehow directly commensurable by giving up the assumption that all outcomes have to be compared on a common scale. In a multi-attribute approach, each type of attribute is measured in the unit the decision maker considers to be most suitable for the attribute in question. Money is

typically the right unit to use for measuring financial costs, whereas other measures are required for measuring car safety.

### **John R. Welch: Real-Life Decisions and Decision Theory**

---

John R. Welch discusses another important area of decision theory, concerning the question of how we can apply decision theory to real-life decisions. As everyone who reads the other chapters on decision theory will quickly discover, decision theorists make many quite unrealistic idealizations about the decision problems they are discussing, which seldom or never hold true in real-life applications. In recent years, decision theorists have become increasingly aware of this limitation, and in response to this have started to develop more realistic decision theories that deal better with the ways in which we actually take decisions in real life. Welch gives a very instructive overview of this emerging literature.

### **Stefan T. Trautmann and Ferdinand M. Vieider: Social Influences on Risk Attitudes: Applications in Economics**

---

In the last chapter on decision theory, Stefan T. Trautmann and Ferdinand M. Vieider discuss the links between decision theory and two of the key disciplines on which much work in decision theory draws, namely, economic theory and psychology. Trautmann and Vieider identify and discuss four distinct types of social influences on economic decisions under risk: (1) observations of other agents' outcomes; (2) observations of the decision maker's outcomes by other agents; (3) direct effects of the decision maker's choices on other agents' outcomes; and (4) direct dependencies of the decision maker's outcomes on other agents' choices.

## **Part 4: Risk Perception**

---

Where the previous section discussed normative or rational decision theory, the present section discusses risk from the point of view of empirical decision theory, or to use a more common notion, risk perception. In the 1970s, the psychologists Amos Tversky, Daniel Kahneman, and Paul Slovic started to investigate the ways in which people as a matter of fact make decisions under uncertainty or risk. It turned out that these decisions deviate significantly from rational decision theory. It might not be too surprising that laypeople's intuitive judgments about risk and statistics deviate from mathematical methods. However, surprisingly, also the judgments of experts turned out to be subject to numerous heuristics and biases. These findings gave rise to a whole research industry in mistakes people make in their risk judgments (cf., e.g., Gilovich et al. 2002), which eventually earned Kahneman the Nobel Prize in economics in 2002. The common framework to explain these phenomena is Dual Process Theory (DPT) which states that there are two systems with which people make judgments: system 1 is intuitive, spontaneous, and evolutionary prior; system 2 is rational, analytical, and comes later in our evolution. System 1 helps us to navigate smoothly through a complex world, but system 2 is the one that provides us with ultimate normative justification.

However, there are alternative approaches to risk perception that challenge this picture to some extent. Paul Slovic and various social scientists have argued that there is no "objective"

measure of risk, that all approaches to risk, also those of experts, involve normative and partially arbitrary or subjective assumptions. Slovic and his colleagues have conducted studies that show that laypeople do not so much have a wrong understanding of risk but rather a different understanding of risk that might provide for valuable insights (cf. Slovic 2000). Dan Kahan has combined Slovic's psychometric approach with Mary Douglas's cultural theory to account for cultural values in risk perception. The psychologist Gerd Gigerenzer (2007) has conducted studies that show that intuitive risk judgments can actually be more reliable than mathematical approaches to risk. For example, experts' intuitions, but also laypeople's heuristics, can be superior to formal approaches.

This section consists of chapters that discuss various aspects of risk perception.

## Dylan Evans: Risk Intelligence

---

Dylan Evans presents new empirical research that shows that risk intelligence perceived as the ability to estimate probabilities correctly is rare. Previous calibration tests have mainly been used to measure expert groups like medics and weather forecasters. However, Dylan presents tests of over 6,000 people of all ages and a variety of backgrounds and countries. Like other work in the psychology of judgment and decision making, Evans' own work shows that most people are not very good at thinking clearly about risky choices. They often disregard probability entirely, and even when they do take probability into account, they make many errors when estimating it. However, some groups of people have an unusually high level of risk intelligence. Evans outlines how lessons can be drawn from these groups to develop new tools to enhance risk intelligence in others.

## Nicolai Bodemer and Wolfgang Gaissmaier: Risk Communication in Health

---

The chapter by Nicolai Bodemer and Wolfgang Gaissmaier studies risk communication in the health-care sector. A major problem for doctors who wish to help their patients to make well-informed medical decisions is that patients often find it difficult to understand the information presented by the doctor. Bodemer and Gaissmaier point out that purely qualitative information, such as saying that the risk is "large" or "quite small," often does not work well: the beliefs that the patients end up with if they receive purely qualitative information is often badly calibrated with the true risk. Numerical representations often make it easier for patients to correctly understand the magnitude of a risk. However, some numerical representations tend to be easier to understand than others. In particular, Bodemer and Gaissmaier argue that doctors should try to avoid using conditional probabilities if they wish to be understood, and instead use natural frequencies.

## Lennart Sjöberg: Risk Perception and Societal Response

---

Lennart Sjöberg discusses research on risk perceptions of experts and laypeople. Sjöberg first reviews the most well-known models of risk perception, that is, the psychometric model by

Paul Slovic and others, and Cultural Theory as developed by Mary Douglas and Aaron Wildavsky (1982). Sjöberg argues that the empirical evidence for these models is problematic. These approaches leave 80% of the variance in risk perception unexplained. Sjöberg presents approaches that have more explanatory power. One tool is the risk sensitivity index with which risk attitudes can be measured. Sjöberg goes on to discuss the role of affect and emotion in risk perception. He points out ambiguities in the use of these notions in studies on risk perception; for example, affect sometimes refers to emotions and sometimes to values. He emphasizes that the psychometric model only employs one emotion, that is, dread, and it is rated for others rather than for the respondent. Sjöberg also discusses social trust, epistemic trust, and antagonism as important dimensions in risk perceptions. He reviews studies that show that risk perceptions of experts and laypeople mainly diverge when experts have responsibilities for risk, which can be explained by self-selection and social validation that lead to lower risk perceptions amongst experts. Sjöberg emphasizes that there will probably never be an ultimate consensus on risk in an open society. He warns for the risks of risk denial.

---

### **Melissa L. Finucane: The Role of Feelings in Perceived Risk**

Melissa L. Finucane discusses the role feelings play in risk perception. She starts with a historical review of the development in research of feeling in risk perception, which is usually placed within the framework of Dual Process Theory. She then discusses different functions of feelings that have been identified in the context of risk, such as providing for accessible information, motivation, and moral and evaluative knowledge. Finucane presents frameworks that question the dichotomies underlying Dual Process Theory and provides for alternative views, for example, that “risk as analysis” (system 2) and “risk as feeling” (system 1) can be combined into what Finucane calls “risk as value.” Finucane goes on by reviewing empirical evidence for the role of feelings in risk perception. One example is the role of emotional images on risk perception. Another example is psychophysical numbing, referring to diminishing sensitivity as numbers of, for example, victims of a disaster increase. This phenomenon can explain why we fail to respond appropriately to large humanitarian or environmental disasters. Finucane also discusses biases in gambles that are due to feelings, and the influence of moods on risk perception. She concludes by pointing out directions for future research, by taking into account available empirical tools for research into feelings, and alternative approaches to risk that go beyond purely quantitative models but also include values and feelings.

---

### **Ross Buck and Rebecca Ferrer: Emotion, Warnings, and the Ethics of Risk Communication**

Ross Buck and Rebecca Ferrer discuss the relation between emotion, warnings, and the ethics of risk communication. They describe the common approach to risk communication which focuses on factual, statistical information. Appeal to emotions is considered to be unethical because it is supposed to be a form of manipulation. Buck and Ferrer challenge this approach. Emotions are already widely used in marketing, often overruling the more sober information on risk, for example, in the context of tobacco and alcohol consumption. They review empirical studies from decision theory and neuropsychology that show the importance of

emotions in decision making. Some of these studies support the framework of Dual Process Theory, others suggest the possibility of an interaction between the affective and the analytical system, and yet others indicate that affect and cognition are intertwined. The authors argue that effective and ethically sound risk communication has to take into account and anticipate the various ways in which emotions can play a role in risk decisions. They present work on the role of emotion in communication about safe sexual behavior and on emotion intervention strategies and emotional education to illustrate this.

### **Dan M. Kahan: Cultural Cognition as a Conception of the Cultural Theory of Risk**

---

Dan M. Kahan discusses two approaches to risk perception: Mary Douglas's and Aaron Wildavsky's cultural theory, and cultural cognition of risk. The latter is a combination of the former with Paul Slovic's psychometric approach to risk. Kahan first provides for a rough outline of cultural theory and its developments. According to cultural theory, cultural worldviews can be fit in a matrix with two axes called "group" and "grid." The group axis ranges from individualism to solidarity, the grid axis from hierarchy to egalitarianism. What distinguishes cultural cognition from other versions of cultural theory is that it allows for a certain way to measure cultural worldviews, a focus on the social and psychological measures that explain the way culture shapes risk perceptions, and a focus on practical applications. The chapter addresses each of these points. It reviews various studies by Kahan and colleagues that show how influential people's cultural worldviews are on the kinds of risks they find salient and which experts they find trustworthy. One way to mitigate that effect is to have experts proclaim unexpected viewpoints, for example, a leftist looking expert making typical right-wing claims. However, Kahan notes that to do this would be ethically dubious. A solution would be to include a plurality of viewpoints in public debates, as this also leads to less predictable views amongst people.

### **Britt-Marie Drottz-Sjöberg: Tools for Risk Communication**

---

Britt-Marie Drottz-Sjöberg studies risk communication projects. She presents different risk communication tasks and derives several general conclusions. Tools for risk communication combine theoretical and applied insights. Drottz-Sjöberg discusses examples which show the influences of social and historic events on a communication setting or a conflict situation and how they shape a risk communication project. She also analyzes how values, attitudes, and feelings influence thinking and behavior within groups. The examples provide for heuristics for the improvement of risk communication. Drottz-Sjöberg also discusses the RISCOM model of transparency. She shows that risk communication always takes place in a social setting, involving various interests, power relations, and actors' own agendas. However, the aim to communicate about specific risks, nevertheless, can be focused on clarification, understanding, and learning. Drottz-Sjöberg describes tools for risk communication that aim at achieving clarity in dialogues, which are characterized by openness and interaction regarding risk issues, in order to enhance problem solving and democracy.

## Part 5: Risk Ethics

---

There is a growing consensus amongst risk scholars that risk is not a purely quantitative notion but also involves qualitative, normative, and ethical considerations. The dominant approach in risk analysis and risk management is to define risk as the probability of unwanted outcomes, such as annual fatalities, and to apply cost–benefit (or risk–benefit) analysis to determine which of various alternative technologies or activities is preferable. However, the question as to which unwanted outcomes to take into account already involves ethical considerations. Furthermore, cost–benefit analysis compares aggregates, whereas it is ethically significant how costs and benefits are distributed within a society. Social scientists and philosophers argue that ethical considerations such as justice, fairness, equity and autonomy have to be taken into account in assessing the acceptability of risk.

Interestingly, the same considerations can be found in the risk perceptions of laypeople that have been studied by Paul Slovic and others and which are discussed in the section on risk perception. Apparently, the intuitive responses to risk by laypeople include ethical aspects. The question arises why these considerations do not figure in the approaches of experts. This might be due to the fact that expert approaches are by definition focused on quantitative data and mathematical tools. Although these approaches can be helpful to a certain degree, they can lead to a tunnel vision that excludes other important considerations. In the context of risk it turns out that laypeople intuitively have a broader perspective that does justice to ethical considerations that can be normatively justified through established ethical theories.

The chapters in this section discuss ethical aspects of risk in more detail.

### Douglas MacLean: Ethics and Risk

---

The point of departure in Douglas McLean's contribution is the widespread belief that traditional ethical theories have little, if anything, to say about risk. Numerous contemporary scholars argue that moral philosophers of the past have simply failed to recognize the ethical issues related to risk, and that this is therefore an area in which more theoretical work is needed. McLean claims that this mainstream picture is not entirely true: Although risk has not been one of the major topics of ethical reflection in the past, it is easy to find examples of scholars who have explicitly discussed ethical principles for risk decisions. The most prominent examples are John Stuart Mill, Jeremy Bentham, and Robert Nozick; the latter, for instance, devoted an entire chapter of his influential book *Anarchy, State, and Utopia* to ethical principles for risk decisions. This means that the ethics of risk has been extensively analyzed within at least two ethical traditions, namely, utilitarianism and theories based on natural rights.

### Carl F. Cranor: Toward a Premarket Approach to Risk Assessment to Protect Children

---

Carl F. Cranor critically discusses risk legislation, specifically the postmarket approach to risk that is currently common in the USA and many other countries. In contrast with this, Cranor argues in favor of a premarket approach on which risks are assessed before products enter the market, similar to legislation concerning pharmaceuticals. Most industrial chemicals are

allowed to enter the market without any testing. Tests are done afterwards, which often leads to strategic behavior, such as claims concerning supposed insufficient evidence about risks. This happened in the case of the tobacco industry, which delayed legislation against smoking for decades. For 70% of the chemicals used for products, there are no toxicity data at all. Cranor reviews evidence of significant amount of traces of dangerous industrial chemicals that can be found in the population. He specifically focuses on health risks for children to give special force to his argument. Given the special vulnerability of small children and fetuses, they should be given additional protection. Legislation should require testing before chemicals are used in consumer products. This will lead to a paradigm shift in legislation as much as in scientific practice.

### **Sabine Roeser: Moral Emotions as Guide to Acceptable Risk**

---

Sabine Roeser explores the role emotions do and can play in debates about risky technologies. Most authors who write on risk and emotion see emotions as a threat to rational decision making about risks. These authors endorse Dual Process Theory, according to which emotion and reason are distinct faculties that have opposite tasks. However, based on recent developments in emotion research, an alternative picture of risk emotions is possible. According to various psychologists and philosophers who study emotions, emotions are a source of practical rationality. They are appraisals or judgments of value that have a cognitive aspect. These ideas can be applied to risk emotions. Emotions such as sympathy and compassion help to grasp morally salient aspects of risk, such as fairness, justice, and autonomy. This view allows for fruitful insights on how to improve public debates about risk, by taking emotional concerns of the public, but also of policy makers and experts, seriously. This approach leads to morally better judgments about risks, by doing justice to emotional-ethical concerns. In addition, as all parties will be taken seriously, it can also help to overcome the gap between experts and laypeople that currently so often leads to a deadlock in discussions about risky technologies.

### **Allison Ross and Nafsika Athanassoulis: Risk and Virtue Ethics**

---

Allison Ross and Nafsika Athanassoulis propose a virtue ethics approach to risk assessment. They argue that it is superior to consequentialist or deontological approaches to the moral assessment of risk. For example, consequentialist approaches to risk are not sensitive to morally salient aspects of risk such as recklessness, fairness, and equity. Risk assessment cannot be left to scientific experts. Risk-taking is both unavoidable and potentially morally problematic. Hence, it requires context-sensitive and reflective judgments. Ross and Athanassoulis argue that it is important to focus on the role of character and patterns of behavior in moral risk assessments. Such patterns should not be understood as the result of arbitrary, automatic processes but as the product of dispositions which constitute somebody's character. Character dispositions are developed through education, habituation, and reflection. They combine desires, emotions, and thoughts that are attuned to decision making about risk in specific circumstances. The authors argue that only virtue ethics with its emphasis on character provides for a framework for sensible and reflective risk judgments. They illustrate this with a hypothetical time-travel experiment in which an agent has to decide about risks for himself or herself and others.

## Philip J. Nickel and Krist Vaesen: Risk and Trust

---

Philip J. Nickel and Krist Vaesen discuss philosophical conceptions of the relationship between risk and trust. They distinguish between three main approaches. The first is a Hobbesian approach. This approach understands trust as a kind of risk assessment about the expected behavior of other people and the estimated benefits of cooperation. This approach comes close to expected utility theory, which is commonly used in formal decision theoretical approaches to risk. The second approach to risk and trust is in direct opposition with such a calculative risk assessment. On this approach, one willingly relies on people based on, for example, habitual, social, or moral reasons. On the third approach, trust is seen as a morally loaded attitude, in which one expects the trusted person to fulfill certain moral obligations. This allows for cooperative behavior in which there are no interpersonal risks. Nickel and Vaesen examine how these three approaches explain relationships between the concepts of risk and trust, also based on empirical research, specifically on cooperative breeding. They suggest that the notion of trust might help overcome the current gap between technocratic and social approaches to risk, if experts are more aware of their moral responsibilities rather than simply providing the public with information, which might create fears.

## Ibo van de Poel and Jessica Nihlén Fahlquist: Risk and Responsibility

---

Ibo van de Poel and Jessica Nihlén Fahlquist discuss the relationship between risk and responsibility. The authors start with noting that even though it is very common to link these two concepts in our daily practice, there is hardly any academic literature on it. They first discuss different conceptions of and connections between risk and responsibility. Van de Poel and Nihlén Fahlquist then elaborate on specific topics concerning responsibility for risk. They discuss the responsibility of engineers to contribute to risk reduction. They elaborate on the role of values and responsibility in risk assessment, risk management, and risk communication. An important distinction is the one between individual and collective responsibility for risk. The authors illustrate this with a case study from traffic safety. Van de Poel and Nihlén Fahlquist suggest various topics for future research, centered around the so-called problem of many hands in relation to climate change. The authors propose three possible lines of research to address this problem, which are responsibility as virtue, procedures for distributing responsibility, and institutional design.

## Madeleine Hayenjelm: What Is a Fair Distribution of Risk?

---

Madeleine Hayenjelm's chapter discusses the question as to what a fair distribution of risk is, partially based on insights from John Rawls's theory of justice. Hayenjelm starts out by reviewing what the objects of fairness are in the context of risk distributions. It is commonsensical that goods should be increased and risks should be diminished. This is also the underlying rationale of risk-cost-benefit analysis. However, such a consequentialist approach to risk overlooks issues of fair distribution, by only focusing on aggregate risks and benefits. Goods and risks should be distributed fairly. This can give rise to moral dilemmas. Hayenjelm discusses problems with equal distributions of probabilities of harm by focusing on a thought

experiment from James Lenman. She then reviews conditions under which deviations of equal distributions are justified and can still be fair. She suggests that this requires the justification of specific risky activities, and that higher risks for specific people should be mitigated by consent, precaution, and compensation.

### **Lauren Hartzell-Nichols: Intergenerational Risks**

---

Lauren Hartzell-Nichols discusses the notion of intergenerational risks – long-term threats of harm that will affect future people – using the example of climate change. She begins by noting that there is comparatively little material on intergenerational risk, and identifies two philosophical problems that are relevant for the issue: Parfit's Non-Identity Problem and Gardiner's Pure Intergenerational Problem. She introduces a distinction between *de re* and *de dicto* badness that can be illuminating. Hartzell-Nichols presents three current approaches to addressing intergenerational risks: (1) cost–benefit analysis, (2) precautionary principles, and (3) approaches based on intergenerational justice, for example, as discussed by Darrel Moellendorf, Henry Shue, Simon Caney, and Steven Vanderheiden. She argues why the two former approaches are problematic. Hartzell-Nichols finally notes that the debate on intergenerational risks point to the larger problem of anthropocentric versus non-anthropocentric ethics.

### **Marko Ahteesuu and Per Sandin: The Precautionary Principle**

---

Marko Ahteesuu and Per Sandin discuss the precautionary principle (PP). The PP is often conceived of as a decision-making principle that calls for early measures to avoid and mitigate hazards in the face of uncertainty, in particular, in the context of environmental problems. Ahteesuu and Sandin trace the PP to three sources: (1) the general idea of precaution, (2) nonjudicial codes of conduct and arguments from precaution, (3) and legal documents. They present three ways of conceiving of the PP: as a rule of choice, a procedural requirement, or as an epistemic principle, and the distinction between weak and strong versions of the PP. Ahteesuu and Sandin also discuss a number of common arguments against the PP, such as that it is ill-defined, self-refuting, or counterproductive. They end with observing that formal methods of inquiry have been insufficiently used in the study of the PP. Some topics that also warrant further research are the normative underpinnings of the principle, the status of the principle in risk analysis, and the relationship between the PP and stakeholder/public engagement.

### **Colleen Murphy and Paolo Gardoni: The Capability Approach in Risk Analysis**

---

Colleen Murphy and Paolo Gardoni discuss the way in which the capability approach might provide for fruitful insights concerning ethical aspects of risk and vice versa. The capability approach has been founded by the economist Amartya Sen and the philosopher Martha Nussbaum and has been extremely influential in the context of development. The capability approach allows to focus on a broader range of capacities, functionings, and achievings than

conventional approaches to development that mainly focus on the availability of goods, but not on what people can do with these goods. The authors discuss three ways in which the capability approach can contribute to risk theory: by focusing on capabilities rather than resources or utility, by focusing on threshold levels of capabilities instead of decision procedures, and by focusing on unacceptable or intolerable risks, which avoids the shortcomings of cost–benefit analysis. On the other hand, there are two ways in which risk theory can contribute to the capability approach, by focusing on security as an important dimension of capability, and by allowing a novel way to assess capabilities that looks further than actual functionings achievements.

## **Part 6: Risk in Society**

---

Given the importance of risk management in modern society, there are several aspects of risk that might, at one point, have seemed to be of mere theoretical interest, but now have vastly important implications for people's lives. The stock and insurance markets rely on methods developed by mathematicians, philosophers, and decision theorists. Increasingly complex technological systems rely on probabilistic methods for safety assessment; methods that could never have been developed without the prior work of risk theorists. In some instances the road from theory to application is short and straight, in other cases long and winding.

As citizens and human beings, we are increasingly required to relate to issues where risk theory is directly relevant in our everyday lives. We are asked to compare insurance policies, invest in the stock market, participate in referendums whether our country should rely on nuclear power or not, and make all kinds of choices where information about likelihood and consequences feed to us from a plethora of different sources.

This has not always been the case. Risk analysis and risk management, as we understand those activities today, are comparatively novel disciplines. We have seen a significant expansion of the field during the last 50 years or so, and the intellectual tools used are modern inventions, where “modern” means at least “post-Renaissance.” The most important of these tools – the mathematical analysis of chance events – is in essence a seventeenth-century invention (or perhaps discovery). It was pioneered by thinkers like Pascal, Descartes, and Bayes, and later refined. Pretty soon, it received its applications in the insurance business. The industrial revolution and the increased scope of the consequences of technology – from steam engines to nuclear power plants to genetically modified organisms and climate change – called for rapid development in the field.

Today, risk consciousness permeates nearly every area of societal life. This has not gone unnoticed by risk theorists, and it has given rise to a number of new disciplines, such as the psychology, sociology, and philosophy of risk. The contributions to this section discuss what Ulrich Beck has famously called “risk society,” or in other words, the ways society does and should cope with risk.

### **Rolf Lidskog and Göran Sundqvist: Sociology of Risk**

---

Rolf Lidskog and Göran Sundqvist begin by reviewing the historical background of general sociology. The focus of sociology is the relationship between society and the individual, and

this holds also for sociology of risk. The authors sketch the history of the sociology of risk and how it started from experts' recognition that public perceptions of risk differed from those of experts, and the attempts to explain this. They then go on to present three central sociological contributions in which risk occupies a prominent place – those of Mary Douglas, Ulrich Beck (1992), and Niklas Luhmann. Then they present five thematic areas which are subject to intense discussion in contemporary sociology of risk: organizational risk, the relation between experts and public, framing and risk, the epistemological status of risk, and governmentality and risk.

### Misse Wester: Risk and Gender: Daredevils and Eco-Angels

---

Misse Wester notes that empirical risk studies show consistent, systematic differences in risk perception between women and men, with focus on environmental issues and disasters. She identifies three different models to explain these differences that have been proposed in the literature: differences in knowledge of and familiarity with science, biological and social differences, and cultural differences. She argues that each model has problems of its own. She hypothesizes that knowledge plays an inferior role in risk perception in comparison to values, ideology, or cultural belonging and calls for further research in this area. She also calls for further research in the form of critical examination of the function of stereotypes in risk issues, and in the form of investigation of empirical studies of how women and men, respectively, are actually affected by crises and risk on a concrete level.

### Tsjalling Swierstra and Hedwig te Molder: Risk and Soft Impacts

---

Tsjalling Swierstra and Hedwig te Molder discuss a bias in current discourse about impact of technology. Policy makers and experts focus on quantifiable and supposedly value-neutral risk, rather than on other, less obviously measurable impacts, such as emotions, values, and subjective experiences. The authors call this “hard” and “soft” impacts, respectively. They examine how this distinction is theoretically and practically construed, by using two case studies. The first case study concerns an online forum for patients with gluten intolerance, and why some patients reject the idea of a pill that might cure them. The reason is that the pill would affect their identity. A second case study concerns how consumers are concerned about the naturalness of food. This concern is dismissed by experts as a private and invalid preference. The authors analyze how social structures shape the distinction between supposedly valid and invalid forms of and concerns about technological impacts. They distinguish how such impacts are evaluated, estimated, and caused. A better understanding of how the demarcation between hard and soft impacts is construed can contribute to overcoming this bias.

### Rinie van Est, Bart Walhout, and Frans Brom: Risk and Technology Assessment

---

Rinie van Est, Bart Walhout, and Frans Brom explore the relationship between risk assessment (RA) and technology assessment (TA), in particular, parliamentary TA, and how that has

evolved over the years since the early days of the Office of Technology Assessment in the USA in the 1970s. The disciplines differ, for instance, with regard to the concept of risk utilized. In RA, risk is typically understood as the product probability and the magnitude of consequences, while TA understands risk in a wider sense, as negative social impact. The authors consider two problems that occur in TA as well as in RA: the problem of representation, that is, who is allowed to define the risk problems under discussion. They discuss how participatory approaches have been developed to alleviate the problem. As a concrete illustration of present-day parliamentary TA, the authors recount the recent TA of nanotechnology carried out by the Rathenau Institute in the Netherlands, and its role in the country's governance of nanotechnology risks.

### **Marijke A. Hermans, Tessa Fox, and Marjolein B. A. van Asselt: Risk Governance**

---

Marijke A. Hermans, Tessa Fox, and Marjolein B. A. van Asselt write about risk governance, by which they mean attempts at an approach to deal responsibly with public risks which is broader in scope than the traditional categories of risk assessment, risk management, and risk communication, and utilizes several different notions of risk in addition to the classical idea of risk as a function of probability and consequences. They review the origins of the approach and the movement from early positivistic approaches to risk. The term “governance” became prominent in the 1980s, originally in studies of development, and was taken over by other subjects. Today the term “governance,” including in risk contexts, is used both in a descriptive and a normative sense, and the distinction is not always clear. Since 2003, considerable efforts have been made by the International Risk Governance Council (IRGC), an independent nonprofit Swiss-based foundation. The authors review current work and analyze it along three lines or principles, which they call “the communication and inclusion principle,” “the integration principle,” and “the reflection principle.”

### **Marjolein B. A. van Asselt and Ellen Vos: EU Risk Regulation and the Uncertainty Challenge**

---

Marjolein B. A. van Asselt and Ellen Vos introduce what they term “the uncertainty paradox,” referring to situations where uncertainty is acknowledged, but where the role of science is seen as providing certainty. The authors argue that it is not recognized that uncertainty undermines the traditional positivist model of knowledge. There are instances of uncertainty intolerance, where uncertainty is not acknowledged and there is unwillingness to produce uncertainty information. To illustrate this, the authors give examples from their analyses of several cases in regulation of risk in the EU, for instance, involving the European Food Safety Authority. They note that uncertainty intolerance is prevalent, but also that there is a tendency to equate uncertainty with risk. They suggest further research involving systematic comparison between risk regulation regimes in different domains. As particularly important topics, they mention the role of science and expertise in decision making and policy, how to deal with uncertainty and trust, the role of the precautionary principle, and stakeholder participation.

## Val Dusek: Risk Management in Technocracy

Dusek critically discusses technocratic risk management approaches. The idea underlying much of the current risk management in Western societies is the assumedly superior expert understanding of risk. Such technocratic tendencies in how we deal with risk can be seen in the large number of government committees, commissions, and corporate departments that issue risk assessments and attempt to manage risks. While several contributions to this handbook challenge the quantitative, Bayesian approach to risk assessment, Dusek critically examines the general attitude of mind underlying this approach. Dusek traces technocratic risk management back to the ideal of the superiority of technocratic rationality as advocated by Plato, the seventeenth-century rationalist as well as Francis Bacon and the British empiricists. He explicates and challenges the ideal of an expert rule, following the technology critique of the critical theory of the Frankfurter Schule and existential philosophy, Dusek does not reject risk analysis and risk management, but thinks that the technocratic trend in risk management which builds on the objectivity, universality, and publicity of science, has to be supplemented by other approaches such as the recent work of Gigerenzer.

## Conclusion

This *Handbook of Risk Theory* unites scholars from disciplines ranging from mathematics and the natural sciences to the social sciences, humanities, and philosophy. However, as diverse as the approaches and topics are, there is one issue that emerges from practically all contributions, namely that risk involves statistics as much as ethics and social values. There are as yet no final answers on how to deal with risk, nor will there probably ever be such answers, but there nevertheless is a consensus that risk should be approached from different perspectives, including those of stakeholders and the public. This requires “sound science” (broadly conceived, i.e., including social sciences, humanities and philosophy), as much as sound political institutions. When it comes to risk, theory and practice are closely intertwined.

## References

- Beck U (1992) Risk society: towards a new modernity. Sage, London
- Douglas M, Wildavsky A (1982) Risk and culture. University of California Press, Berkeley
- Gigerenzer G (2007) Gut feelings: the intelligence of the unconscious. Viking, London
- Gilovich T, Griffin D, Kahnemann D (eds) (2002) Intuitive judgment: heuristics and biases. Cambridge University Press, Cambridge
- Hume D (1740) A treatise of human nature, 1967th edn. Oxford University Press, Oxford
- Ramsey FP (1931) Truth and probability. In: The foundations of mathematics and other logical essays. Routledge and Kegan Paul, London, 156–198
- Rawls J (1971) A theory of justice. Harvard University Press, Cambridge
- Savage LJ (1954) The foundations of statistics. Wiley, New York
- Slovic P (2000) The perception of risk. Earthscan, London
- von Neumann J, Morgenstern O (1944) Theory of games and economic behavior. Princeton University Press, Princeton



## **General Issues in Risk Theory**



# 2 A Panorama of the Philosophy of Risk

Sven Ove Hansson

Royal Institute of Technology, Stockholm, Sweden

<i>Introduction</i> .....	28
<i>Historical Background</i> .....	28
<i>What Philosophy Can Contribute</i> .....	29
Terminological Clarification .....	30
Argumentation Analysis .....	31
The Fact-Value Distinction .....	32
<i>Philosophical Perspectives in Risk Theory</i> .....	33
Epistemology .....	34
The Limits of Epistemic Credibility .....	34
The Legitimacy of Expertise in Uncertain Issues .....	35
Decision Theory .....	35
Philosophy of Probability .....	36
Philosophy of Science .....	37
Philosophy of Technology .....	40
Inherent Safety .....	40
Safety Factors .....	41
Multiple Independent Safety Barriers .....	41
Ethics .....	43
Philosophy of Economics .....	48
Political Philosophy .....	50
<i>Further Research</i> .....	51

**Abstract:** The role of philosophy in the development of the risk sciences has been rather limited. This is unfortunate since there are many problems in the analysis and management of risk that philosophers can contribute to solving. Several of the central terms, including “risk” itself, are still in need of terminological clarification. Much of the argumentation in risk issues is unclear and in need of argumentation analysis. There is also still a need to uncover implicit or “hidden” values in allegedly value-free risk assessments. Eight philosophical perspectives in risk theory are outlined: From the viewpoint of *epistemology*, risk issues have brought forth problems of trust in expertise and division of epistemological labor. In *decision theory*, the decision-maker’s degree of control over risks is often problematic and difficult to model. In the *philosophy of probability*, posterior revisions of risk estimates (in so-called hindsight bias) pose a challenge to the standard model of probabilistic reasoning. In the *philosophy of science*, issues of risk give us reason to investigate what influence the practical uses of knowledge can legitimately have on the scientific process. In the *philosophy of technology*, the nature of safety engineering principles and their relationship to risk assessment need to be investigated. In *ethics*, the most pressing problem is how standard ethical theories can be extended or adjusted to cope with the ethics of risk taking. In the *philosophy of economics*, the comparison and aggregation of risks falling to different persons give rise to new foundational problems for the theory of welfare. In *political philosophy*, issues such as trust and consent that have been discussed in connection with risk give us reason to reconsider central issues in the theory of democracy.

## Introduction

---

Philosophy is often seen as an unworldly discipline, dealing with abstract and contrived issues that have very little connection with real life. Concededly, philosophy has a long tradition of unabashedly delving into intellectual problems that have no immediate application. In this it does not differ from most other academic disciplines. But philosophy also has another side. It has a strong tradition, going back at least to Socrates and Aristotle, of probing into issues that societies and individuals need to understand better in order to solve practical problems. And just as in other disciplines, some of the progress made in studies driven by pure intellectual curiosity has turned out to provide us with indispensable tools for investigations aimed at solving practical problems. Examples of this can be found in the philosophy of risk as well as other areas of applied philosophy.

In what follows, a brief historical background (section [❶ Historical Background](#)) will be followed by a presentation of the major types of contributions that philosophy can make to risk research (section [❷ What Philosophy Can Contribute](#)) and an overview over eight philosophical perspectives on risk (section [❸ Philosophical Perspectives in Risk Theory](#)). Finally, some topics for further research will be summarized (section [❹ Further Research](#)).

## Historical Background

---

Modern risk research originated in studies from the 1960s and 1970s that had a strong focus on chemical risks and the risks associated with nuclear energy. From its beginnings, risk research drew on competence in areas such as toxicology, epidemiology, radiation biology, and nuclear engineering. Today, many if not most scientific disciplines provide risk analysts with specialized

knowledge needed in the study of one or other type of risk – medical specialties are needed in the study of risks from diseases, engineering specialties in studies of technological failures, etc. In addition, several disciplines have supplied overarching approaches to risk, intended to be applicable to risks of different kinds. Statistics, epidemiology, economics, psychology, anthropology, and sociology are among the disciplines that have developed general approaches to risk.

Philosophers did not have a big role in the early development of risk analysis. Most of the philosophical contributions to the area were in fact outsiders' criticisms of risk analysis. There was a strong tendency in the early development of risk analysis to downplay value issues. Risk assessments were presented as objective scientific statements, even when taking a stand on value-laden issues such as risk acceptability. Most of the early philosophical work on risk had as its main purpose to expose the value-dependence of allegedly value-free risk assessments (Thomson 1985b; MacLean 1985; Shrader-Frechette 1991; Cranor 1997; Hansson 1998). This was an important task, and it was also undertaken with some success. Although hidden value assumptions are still common in risk assessments, there is now much more awareness of their presence. Philosophers who took part in these discussions certainly contributed to the steps that have been taken to keep facts and values apart as far as possible in the assessment of risk, in particular attempts to divide the risk-decision process into a fact-finding risk-assessment part and a value-based risk-management phase (National Research Council 1983).

In the 1990s, philosophers increasingly discovered many other risk-related issues in need of philosophical clarification. Philosophers have studied the nature of risks, the specific characteristics of knowledge about risk, the ethics of risk taking, its decision-theoretical aspects, the implications of risk in political philosophy, and several other areas. It is too early to write the history of these developments, but a pattern emerges in which most of the major subdisciplines of philosophy turn out to have important risk-related issues to deal with. These developments will be introduced in section  [Philosophical Perspectives in Risk Theory](#), but before that we are going to look more closely at the nature of the philosophical contribution.

## What Philosophy Can Contribute

---

Philosophy is unique in having potential connections with virtually every other academic discipline. Philosophical concepts and methods have proven to be applicable to a wide variety of problems in other academic disciplines. When you probe into almost any field of learning, interesting problems of a philosophical nature tend to emerge. Unfortunately, this potential is underused, largely due to intellectual isolation and to the “two-cultures” phenomenon that separates philosophers from empirical scientists.

The contributions of philosophy to other disciplines and to interdisciplinary cooperations can be of many kinds, but experience shows that there are certain ways in which philosophy has particularly often turned out to be useful. Three of them are especially important in risk research:

- *Terminological clarifications:* Philosophy has a long tradition of constructing precise definitions and developing new distinctions, often beyond the limits of what can readily be expressed with current linguistic means. Armed with standard tools and distinctions from philosophy, philosophers can often contribute to conceptual clarification in other disciplines.

- *Argumentation analysis:* Arguments, as we express them in scientific or social debates, tend to depend on unstated assumptions. Using the tools of logic and conceptual analysis philosophers can often exhibit hidden assumptions and clarify the structure of arguments.
- *The fact–value distinction:* Factual input from science has a large and increasing role in debates on social issues. This applies to virtually all branches of science: economics, behavioral science, environmental science, climatology, medical science, technological sciences, etc. But even if scientists try to make their statements as value-independent as possible, they do not always succeed in this. Philosophical tools are useful in identifying the values that are inherent in science-based information.

In the following three subsections, we will have a brief look at each of these types of contribution, in order to show how philosophical method can contribute to investigations of risk that are performed primarily by researchers in other fields.

## Terminological Clarification

---

As in many other research areas, the terminology in risk research is often imprecise. This applies even to key terms such as “risk” and “safety.” The word “risk” has been taken over from everyday language, where it is used (often somewhat vaguely) to describe a situation in which we do not know whether or not some undesired event will occur. In risk analysis, two major attempts have been made to redefine risk as a numerical quantity. First, in the early 1980s, attempts were made to identify risk with the probability of an unwanted event (Fischhoff et al. 1981; Royal Society 1983, p. 22). This usage has some precedents in colloquial language; we may for instance say that “the risk that this will happen is one in twenty.” Secondly, in more recent years, several attempts have been made to identify risk with the statistical expectation value of unwanted events. By this is meant the product of an event’s probability with some measure of its undesirability. If there is a probability of 1 in 100 that three people will die, then “the risk” is said to be 0.03 deaths. Currently, this is by far the most common technical definition of risk (International Organization for Standardization 2002; Cohen 2003).

From the viewpoint of philosophical definition theory (Hansson 2006b), this terminology is problematic in at least two ways. First, it conflates “risk” with “severity of risk.” It makes sense to say that a probability of 1 in 1,000 that one person will die in a roller coaster accident is “equally serious” as a probability of 1 in 1,000,000 that 1,000 people will die in a nuclear accident, but it does not make sense to say that these two are the same risk (namely 0.001 deaths). They are in fact risks with quite different characteristics.

Secondly, it is a controversial value statement that risks with the same expectation value of undesirable events are always equally serious. Some authors have claimed that serious events with low probabilities should be given a higher weight in decision making than what they receive in the expected utility model (O’Riordan and Cameron 1994; O’Riordan et al. 2001; Burgos and Defeo 2004). The identification of “risk” with expectation values has the unfortunate effect of ruling out this view on the severity of risk by means of a terminological choice. In order to achieve clarity in discussions on risk, we need to make a clear distinction between a risk and its severity, and we also need to avoid terminology that takes controversial standpoints on what constitutes risk severity for granted. Therefore, a term such as “expected damage” is much preferable to “risk” as a designation of the statistical expectation values employed in risk analysis (Hansson 2005).

Besides “risk,” several other terms used in risk studies are in need of terminological clarification. Prominent among these are “safety” and “precautionary principle.”

“Safety” has sometimes been defined as a situation without accidents (Tench 1985) and on other occasions as a situation with an acceptable probability of accidents (Miller 1988). In a recent philosophical analysis of the concept, it was shown that usage of the terms “safe” and “safety” vacillates between an absolute concept (“safety means no harm”), and a relative concept that only requires such risk reductions that are considered to be feasible and reasonable. It may not be possible to eliminate either of these usages, but it is possible to keep track of them and avoid confusing them with each other (Möller et al. 2006).

The “precautionary principle” is a principle for decision making under scientific uncertainty that has been codified in a several international treaties on environmental policies. Its major message is that policy decisions in environmental decisions can legitimately be based on scientific evidence of a danger, even if that evidence is not strong enough to constitute full scientific proof that the danger exists. There has been considerable controversy on the precise meaning of the principle. A careful philosophical analysis showed that the major definitions of the precautionary principle contain four major components, namely (1) a threat to the environment or to human health, (2) a degree of uncertainty that is sufficient for action (such as “even before scientific proof is established”), (3) the action that is then taken (e.g., “warn” or “forbid”), and (4) the level of prescription (e.g., “is mandatory”) (Sandin 1999). The first two of these can be summarized as the *trigger* of the precautionary principle, whereas the last two constitute the *precautionary response* (Ahteesuu 2008). Although this analysis does not resolve the controversies on the principle, it facilitates a precise understanding of these controversies.

## Argumentation Analysis

---

Ever since Aristotle, logical and argumentative fallacies have been an important topic in philosophy (Walton 1987). It is not difficult to find examples of traditional fallacies such as ad hominem in discussions on risk. In addition, there are fallacies that are specific to the subject matter of risk. The following is a sample of such fallacies:

Risk X is accepted.

*Y is a smaller risk than X.*

∴ Y should be accepted.

*Risk X is natural.*

∴ X should be accepted.

*X does not give rise to any detectable risk.*

∴ X does not give rise to any unacceptable risk.

*There is no scientific proof that X is dangerous.*

∴ No action should be taken against X.

*Experts and the public do not have the same attitude to risk X.*

∴ The public is wrong about risk X.

*A's attitude to risk X is emotional.*

∴ A's attitude to risk X is irrational.

For examples of the first five of these fallacies and clarifications of why they are fallacies, see Hansson (2004b). For the last of these fallacies, see Roeser (2006).

## The Fact-Value Distinction

---

As already mentioned, the task of uncovering hidden value assumptions in risk assessments often requires philosophical competence. Implicit value components of complex arguments have to be discovered, and conceptual distinctions relating to values have to be made. Often, other competences are required as well. A thorough understanding of the technical contents of risk assessments is needed in order to determine what factors influence their outcomes (Hansson and Rudén 2006). This is an area in which cooperations between philosophy and other disciplines can be very fruitful.

For the philosophical part of this work, two distinctions are particularly important. The first is the seemingly trivial but, in practice, often overlooked distinction between being value-free and being free of controversial values. There are many values that are shared by virtually everyone or by everyone who takes part in a particular discourse. Medical science provides good examples of this. When discussing analgesics, we take for granted that it is better if patients have less rather than more pain. There is no need to interrupt a medical discussion in order to point out that a statement that one analgesic is better than another depends on this value assumption. Similarly, in economics, it is usually taken for granted that it is better if we all become richer. Economists sometimes lose sight of the fact that this is a value judgment. Obviously, a value that is uncontroversial in some circles may be controversial in others. This is one of the reasons why values believed to be uncontroversial should be made explicit and not treated as non-values.

The other distinction is that between epistemic and non-epistemic values. Most of the values that we usually think of in connection with risk policies are non-epistemic. The epistemic values are those that rule the conduct of science. Among the most commonly mentioned examples of such values are the attainment of truth, the avoidance of error, simplicity, and explanatory power. It was Carl Hempel who pointed out that these should be treated as values, although they are not moral values (Hempel 1960; Levi 1962; Feleppa 1981; Harsanyi 1983). Epistemic values are not necessarily less controversial than non-epistemic ones, but these are different types of controversies that should be kept apart.

The following are three examples of values that are often implicit or “hidden” in risk assessments:

1. *Values of error-avoidance:* Two major types of errors can be made in a scientific statement. Either you conclude that there is a phenomenon or an effect that is in fact not there. This is called an error of type I (a false positive). Or you miss an existing phenomenon or effect. This is called an error of type II (a false negative). In scientific practice, errors of type I are the more serious ones since they make us draw unwarranted conclusions, whereas errors of type II only make us keep an issue open instead of adopting a correct hypothesis. As long as we stay in the realm of pure science, the relative weights that we assign to the two types of error express our epistemic values, and they need not have any connection with our non-epistemic values. However, when scientific information is transferred to risk assessment, values of error-avoidance are transformed into non-epistemic and often quite controversial values. Consider the question “Does Bisphenol A impair infant brain development?” In a purely scientific context, the level of evidence needed for an affirmative answer to this question is a matter of

epistemic values. (How close to certainty should we be in order to take something to be a scientific fact?) In a risk assessment context, the relevant issue is what level of evidence we need to act as if the substance has this effect. This is a matter of non-epistemic values. (How much evidence is needed for treating the substance as toxic to infants?) In a case like this, a focus on epistemic values will usually lead to more weight being put on the avoidance of type I errors than type II errors, whereas a focus on non-epistemic values can have the opposite effect.

The distinction between a scientific assessment and a judgment of what should be done given the available scientific information is both fundamental and elementary, but it is nevertheless often overlooked (Rudén and Hansson 2008). It is not uncommon to find scientists unreflectingly applying epistemic standards of proof in risk assessment contexts where they are not warranted. It should be said to their defense that this is often more difficult to avoid than what one would perhaps think. Scientists are educated to focus on type I errors, and the tools of science are often ill suited to deal with type II errors. As one example of this, standard statistical practices for the evaluation of empirical data that have been tailored to the epistemic issue need to be adjusted in order to deal adequately with the risk assessment situation in which type II errors are usually more important (Krewski et al. 1989; Cranor and Nutting 1990; Leisenring and Ryan 1992; Hansson 1995, 2002).

2. *The value of naturalness:* In public debates, risks associated with GMOs or synthetic chemicals are often denounced as “unnatural.” This argument is seldom used in risk assessments, but a converse version of it can sometimes be found, most often in connection to radiation. Radiation levels are frequently compared to the natural background with the tacit assumption that exposures lower than the natural background are unproblematic. In health risk assessments, this is a very weak argument. That something is natural does not prove that its negative effects on human health are small (Hansson 2003a). In ecological risk assessments, an argument referring to naturalness may be more relevant. If we want to protect the natural environment, then it is important to know what is natural. However, appeals to naturalness or unnaturalness are often made in a perfunctory way in discussions on ecological risk, and there is much need for clarification and analysis.
3. *Attitudes to sensitive individuals:* Risk assessments tend to focus on individuals with average sensitivity to the exposure in question. However, individual sensitivity differs and in many cases it is possible to identify groups of exposed persons who run a larger risk than others. According to the best available estimates, the radiogenic cancer risk is around 40% higher for women than for men at any given level of exposure. There are also small groups in the population who run a much higher risk. However, the recommended exposure limits are based on a population average rather than data for subpopulations (Hansson 2009a). From an ethical point of view, this is problematic. Exposing a person to a high risk cannot be justified by pointing out that the risk to an average person would have been much lower. Nevertheless, sensitive groups are often overlooked or disregarded in risk assessments, and the ethical implications of doing so are seldom discussed. It often takes careful study to reconstruct and analyze the underlying value assumptions.

## Philosophical Perspectives in Risk Theory

In the previous section, we encountered several ways in which philosophers can contribute to interdisciplinary risk studies. Such contributions can be seen as applications of philosophy,

mostly without much influence on the core of philosophical research. But the interaction between philosophy and risk studies does not end there. In recent years, it has become increasingly clear that risk has implications in many if not most of the philosophical subdisciplines. In what follows, we will have a look at several of these subdisciplines. In some of them, there is already an established tradition of studying risk. In others, little has yet been done, but interesting issues for future research can nevertheless be pointed out.

## Epistemology

---

Risks are always connected to lack of knowledge. If we know for certain that there will be an explosion in a factory, then there is no reason for us to talk about that explosion as a risk. Similarly, if we know that no explosion will take place, then there is no reason either to talk about risk. What we refer to as a risk of an explosion is a situation in which it is not known whether or not an explosion will take place. In this sense, knowledge about risk is knowledge about the unknown. It is therefore a quite problematic type of knowledge. It gives rise to several important epistemological questions that have not been much studied. Two of them will be mentioned here.

### The Limits of Epistemic Credibility

Some issues of risk refer to possible dangers that we know very little about. Recent debates on biotechnology and nanotechnology are examples of this. It is easy to find examples in which many of us would be swayed by considerations of unknown dangers. Suppose that someone proposed to eject a chemical substance into the stratosphere in order to compensate for the anthropogenic greenhouse effect. It would not be irrational to oppose this proposal solely on the ground that it may have unforeseeable consequences, even if all specified worries can be neutralized.

But on the other hand, it would not be feasible to take the possibility of unknown effects into account in all decisions that we make. Given the unpredictable nature of actual causation, almost any decision may lead to a disaster. We therefore have to disregard many of the more remote possibilities. It is easy to find examples in which it can be seen in retrospect that it was wise to do so. In 1969, *Nature* printed a letter that warned against producing polywater, polymerized water. The substance might “grow at the expense of normal water under any conditions found in the environment,” thus replacing all natural water on earth and destroying all life on this planet (Donahoe 1969). Soon afterward, it was shown that polywater does not exist. If the warning had been heeded, then no attempts would have been made to replicate the polywater experiments, and we might still not have known that polywater does not exist. In cases like this, appeals to the possibility of unknown dangers may stop investigations and thus prevent scientific and technological progress.

It appears to be an unavoidable conclusion that we should take some but not all remote possibilities seriously. But which of them? What about the warnings that global warming might soon be aggravated by feedbacks that lead to a run-away greenhouse effect totally beyond our control, the warnings that the greenhouse effect may not exist at all, the warnings that mobile phones might have grave health effects, and that high-energy physics experiments might lead to an apocalypse? We are in need of concepts and criteria to discuss such issues in a systematic way, but as yet very little research has been performed on how to assess epistemic credibility in cases like this (Hansson 1996, 2004d).

## The Legitimacy of Expertise in Uncertain Issues

In many issues of risk, we have seen wide divergences between the views of experts and those of the public. This is clearly a sign of failure in the social system for division of intellectual labor. However, it should not be taken for granted that every such failure is located within the minds of the nonexperts who distrust the experts. Experts are known to have made mistakes. A rational decision-maker should take into account the possibility that this may happen again. This will be particularly important in cases when experts assign very low probabilities to a highly undesirable event. Suppose that a group of experts have studied the possibility that a new microorganism that has been developed for therapeutic purposes will mutate and become virulent. They have concluded that the probability that this will happen is 1 in 100,000,000. Decision-makers who receive this report should of course consider whether this is an acceptable probability of such an event, given the advantages of using the new organism. But, arguably, this is not the most important question they should ask. The crucial issue is how much reliance they should put on the estimate. If there is even a very small probability that the experts are wrong, say a probability that we in some way estimate as 1 in a million, then that will be the main problem to deal with. In cases like this, reliance on experts creates serious epistemic problems that we do not yet seem to have adequate tools to analyze.

## Decision Theory

---

A risk (in the informal sense of the word) is a situation in which some undesirable event may or may not occur, and we do not know which. Probability theory is a tool for modeling such situations. However, it should not be taken for granted that all such situations can be adequately modeled in that way. In many cases, our knowledge is so incomplete that no meaningful probability estimates are obtainable. In other cases, the situation may have features that make it unsuitable for probabilistic modeling. This applies in particular to risks that depend on complex interactions between independent agents. We all try both to influence the choices that others make and to foresee them and adjust to them. Therefore, our choices will depend in part on how we expect others to react and behave, and conversely their choices will depend on what they expect from us. Such interpersonal interactions are extremely difficult to capture in probabilistic terms.

This applies not least to malevolent action, such as the actions of an enemy, a saboteur, or a terrorist. Such agents try to take their adversaries with surprise. It is in practice impossible – and perhaps even counterproductive – to make probability estimates of their actions. For most purposes, a game-theoretical approach that makes no use of probabilities is more adequate to deal with inimical actions than models that employ probability estimates.

The use of probabilistic models is also problematic in situations where we have to take a whole series of decisions into account. The crucial issue here is whether or not one should treat one's own future choices and decisions as under one's present control (Spohn 1977; Rabinowicz 2002). The consequences at time  $t_3$  of your actions at time  $t_1$  are not determinate if you have an intermediate decision point  $t_2$  at which you can influence what happens at  $t_3$ . In a moral appraisal of your actions at  $t_1$ , you have to decide whether to treat your actions at  $t_2$  as under your own control at  $t_1$  or as beyond your control at that point in time. In the former case, a decision at  $t_1$  can bind your actions at  $t_2$ ; in the latter case, it cannot do so. Consider the following two examples:

### Example 1

A nonsmoker considers the possibility to smoke for just 1 week and then stop in order to achieve a better understanding of why so many people smoke. When making this decision, should she regard herself as being in control of the future decision whether or not to stop after a week? Or should she make a probabilistic appraisal of what she will do in that situation?

### Example 2

A heavy cigarette smoker considers whether or not to try to quit. When making this decision, should she regard herself as being in control of future decisions whether or not to start smoking again? Or should she make a probabilistic appraisal of her future decisions? From such a viewpoint, quitting may seem to have a too meager prospect of success to be worth trying (Hansson 2007a).

Probably, most of us would recommend the non-control (probabilistic) approach to future decisions in Example 1 and the control (non-probabilistic) approach in Example 2. However, no general rule seems to be available to determine when a probabilistic approach to one's own future decisions is appropriate. Since most risk issues seem to require decisions on more than one occasion, this is a problem with high practical relevance. We have access to sophisticated decision-theoretical models that employ probabilities, but we do not have tools to determine when we should use these models and when we should instead use non-probabilistic approaches.

---

## Philosophy of Probability

An average person's yearly risk of being struck by lightning is somewhat below one in a million (Lopez and Holle 1998). Risks of that magnitude have often been considered "negligible." After the Deepwater Horizon oil spill in 2010, BP's chief executive, Tony Hayward, said that the risk of this spill had been "one in a million" (Cox and Winkler 2010). However, there was no sign that the public – or the public authorities – were willing to discuss BP's responsibility from the premise that the accident was almost as unlikely as being struck by lightning. The fact that the accident had occurred was generally taken as proof that the company had not taken sufficient measures to prevent it from happening.

This example reveals a common pattern in how we argue about probabilities in the context of risks. If the expected utility argumentation were followed to the end, then many accidents would be defended as consequences of a maximization of expected utility that is, *in toto*, beneficial. But, such an argument is very rarely heard in practice. Once a serious accident has happened, not much credence is given to the calculations showing that it was highly improbable. Instead, the very fact that the accident happened is taken as evidence that its probability was higher than estimated. Such reasoning has been disparaged by some as a fallacy, "hindsight bias" (Levi 1973; Fischhoff 1977). But it is not a fallacy.

Suppose that we know that a certain accident took place yesterday. Then the probability that it *did* happen was 1. Nevertheless, the probability that it *would* happen can have been much lower, say 1 in 100. As was observed by Blackburn in a different context, these are two distinctly different types of probabilities. "We can say that the probability of an event was high at some

time previous to its occurrence or failure to occur, and this is not to say that it is now probable that it did happen" (Blackburn 1973, p. 102).

A simple example serves to show that our estimates of past probabilities can legitimately be influenced by information about what happened after the events in question. Suppose that a die was tossed 1,000 times yesterday. My original belief about the chance of a six on the first toss was that it was 1/6. When I learn that all the 1,000 tosses yielded a six, I change my opinion and assign a probability close to 1 to the event to which I previously had assigned 1/6. This is sensible since it is much more plausible that a die is biased than that a fair die yields the same outcome in 1,000 tosses.

The same principle applies to the probabilities referred to in risk analysis, such as the probability of an accident. Suppose that a new type of nuclear reactor is built. Nuclear engineers argue persuasively that it is much safer than previous designs. They convince us that the probability of a core damage ("meltdown") is 1 in  $10^8$ . However, after the first reactor of the new type has been in service for only a couple of months, a serious accident involving core damage occurs. Probably, most people would not see this as an example of an extremely improbable event taking place. Instead, they would see the accident as a very strong indication that the probability estimate 1 in  $10^8$  was wrong. Just as in the example with the die, it would be perfectly rational to substantially revise one's estimate of the probability, perhaps from 1 in  $10^8$  to 1 in  $10^5$  or even higher. However, this is not an ordinary (Bayesian) revision of probabilities (A Bayesian revision refers to the probability that the accident actually took place, and thus takes us all the way from  $10^{-8}$  to 1). This is a nonstandard form of probabilistic revision. It can be accounted for in terms of second-order probabilities (Hansson 2009b, 2010b), but its properties and its implications in assessments of risk remain to be investigated.

## Philosophy of Science

In order to understand the relationship between risk assessments and scientific knowledge, it is useful to take intrascientific knowledge production as a starting point. The production of scientific knowledge begins with data that originate in experiments and other observations. Through a process of critical assessment, these data give rise to the scientific corpus (See [Fig. 2.1](#)). The corpus consists of that which is taken for given by the collective of researchers in their continued research and, thus, not questioned unless new data give reason to question it (Hansson 2007b). Hypotheses are included into the corpus when the data provide sufficient evidence for them, and the same applies to corroborated generalizations that are based on explorative research.



**Fig. 2.1**  
The knowledge-formation process in pure science

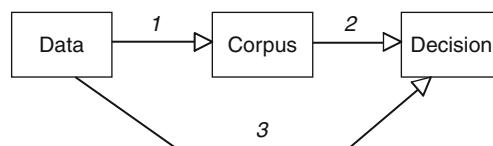
The scientific corpus is a highly complex construction, much too large to be mastered by a single person. Different parts of it are maintained by different groups of scientific experts. These parts are all constantly in development. New statements are added, and old ones removed, in each of the many subdisciplines, and a consolidating process based on contacts and cooperations among interconnected disciplines takes place continuously. In spite of its complex structure, the corpus is, at each point in time, reasonably well defined. In most disciplines, it is fairly easy to distinguish those statements that are, for the time being, generally accepted by the relevant experts from those that are contested, under investigation, or rejected. Hence, although the corpus is not perfectly well defined, its vague margins are fairly narrow.

The process that leads to modifications of the corpus is based on strict standards of evidence that are an essential part of the ethos of science. Those who claim the existence of an as yet unproven phenomenon have the burden of proof. In other words, the corpus has high entry requirements. This is essential to protect us against the importation of false beliefs into science.

But as we noted in section [❶ The Fact – Value Distinction](#), scientific information is often used not only to guide the progress of science but also to guide practical decisions. As one example of this, studies of the anthropogenic greenhouse effect are used both to achieve more reliable scientific knowledge about what happens to the climate and to determine what practical decisions to take in climate policies. In this and many other cases, two decisions have to be based on the same scientific information: the intrascientific decision concerning what to believe and an extrascientific (practical) decision concerning what to do. These are two different decisions, although they make use of the same scientific data.

[❷ Figure 2.2](#) illustrates the practical use of scientific information (Hansson 2004c). The obvious way to use science for decision-guiding purposes is to employ information from the corpus (arrow 2). In many cases, this is all that we need to do. The high entrance requirements of the corpus have the important effect that the information contained in it is dependable enough to be relied on in almost all practical contexts. Only on very rare occasions do we need, for some practical purpose, to apply stricter standards of evidence than those that regulate corpus inclusion.

However, the high entry requirements of the corpus also have another, more complicating implication. On some occasions, evidence that was not strong enough for corpus entry may nevertheless be strong enough to have legitimate influence in some practical matters. To exemplify this, suppose that a preservative agent in baby food is suspected of having a negative health effect. The evidence weighs somewhat in the direction of there being an effect, and most scientists consider it to be more probable that the effect exists than that it does not. Nevertheless, the evidence is not conclusive, and the issue is still open from a scientific point of view. Considering what is at stake, it would be perfectly rational for a food company or a government agency to cease the use of the substance. Such a decision would have to be informed by scientific information that did not satisfy the criteria for corpus entry. More generally speaking, it would not seem rational – let alone morally defensible – for a decision-maker



**Fig. 2.2**

The use of scientific data for decision-making

to ignore all preliminary indications of a possible danger that do not amount to full scientific proof. We typically wish to protect ourselves against suspected health hazards even if the evidence is much weaker than what is required for scientific proof. As was indicated in section **● The Fact – Value Distinction**, in order to guide the type of decisions that we want to make, these decisions have to be based on standards of evidence that differ from the criteria used for intrascientific purposes. Evidence that is weaker than the requirements for corpus entry cannot influence decisions in the “standard” way that is represented in **● Fig. 2.2** by arrows 1 and 2. In cases like this, we need to take a direct way from data to practical decision-making (arrow 3).

Just like the process represented by arrow 1, the bypass route represented by arrow 3 involves an evaluation of data against criteria of evidence. However, the two evaluation processes differ in being calibrated to different criteria for the required strength of evidence. The process of arrow 1 is calibrated to the standard scientific requirements, whereas that of arrow 3 is calibrated to criteria corresponding to the needs of a practical decision. However, the latter process is nevertheless in important respects a scientific one. From the viewpoint of philosophy of science, it is a challenge to clarify the nature of argumentation and decision processes like this that contain a mixture of scientific and policy-related components.

From a somewhat more practical point of view, it is essential to ensure that the bypass route does not lead to inefficient use of the available scientific information. In order to see what this requires, it is instructive to compare the processes represented by arrows 1 and 3. First of all, there should be no difference in the type of evidence that is taken into account. Hence, in the baby-food example, the same experimental and epidemiological studies are relevant for the intrascientific decision (arrow 1) and for the practical one (arrow 3). The evidence is the same, although it is used differently. Furthermore, the assessment of how strong the evidence is should be the same in the two processes. What differs is the *required* level of evidence for the respective purposes (Hansson 2008).

The term “precautionary principle” has often been used to designate the process illustrated by arrow 3 in our diagram (cf. section **● Terminological Clarification**). But the need for a special principle can be put in doubt. Once it is recognized that the principle applies to practical decisions, it will be seen that the importation of practical values that this route makes possible is not only legitimate but in many cases also rationally required. From a decision-theoretical point of view, allowing decisions to be influenced by uncertain information is not a special principle that needs to be specially defended. To the contrary, doing so is nothing else than ordinary practical rationality, as it is applied in most other contexts. If there are strong scientific indications that a volcano may erupt in the next few days, decision-makers will expectedly evacuate its surroundings as soon as possible, rather than waiting for full scientific evidence that the eruption will take place. More generally speaking, it is compatible with – and arguably required by – practical rationality that decisions be based on the available evidence even if it is incomplete.

Although the account given here is a reasonable ideal account, it is far from easy to implement in practice. If we want to take uncertain indications of toxicity seriously, then this has implications not only on how we interpret toxicological tests but also on our appraisals of more basic biological phenomena. If our main concern is not to miss any possible mechanism for toxicity, then we must pay serious attention to possible metabolic pathways for which there is insufficient proof. Such considerations in turn have intricate connections with various issues in biochemistry and, ultimately, we are driven to reappraise an immense number of empirical conclusions, hypotheses, and theories. Due to our cognitive limitations, this cannot

in practice be done. In practice, we will have to rely on the corpus in most issues and use the detour (arrow 3) only in a limited number of selected issues. It remains to clarify how such partial adjustments are best made.

## Philosophy of Technology

---

Since the nineteenth century, engineers have specialized in worker's safety and other safety-related tasks. With the development of technological science, the ideas behind safety engineering have been subject to academic treatments. However, most of the discussion on safety engineering is fragmented between different areas of technology. The same basic ideas or "safety philosophies" seem to have been developed more or less independently in different areas of engineering. Therefore, the same or similar ideas are often discussed under the different names for instance by chemical, nuclear, and electrical engineers. But a recent study has shown that there is much unity in this diversity. In spite of the terminological pluralism, and the almost bewildering number of similar or overlapping safety principles, much of the basic thinking seems to be the same in the different areas of safety engineering (Möller and Hansson 2008). In order to see what these basic ideas are, let us consider three major principles of safety engineering: inherent safety, safety factors, and multiple barriers.

### Inherent Safety

Also called primary prevention, inherent safety consists in the elimination of a hazard. It is contrasted with secondary prevention that consists in reducing the risk associated with a hazard. For a simple example, consider a process in which inflammable materials are used. Inherent safety would consist in replacing them by noninflammable materials. Secondary prevention would consist in removing or isolating sources of ignition and/or installing fire-extinguishing equipment. As this example shows, secondary prevention usually involves added-on safety equipment.

The major reason to prefer inherent safety to secondary prevention is that as long as the hazard still exists, it can be realized by some unanticipated triggering event. Even with the best of control measures, if inflammable materials are present, some unforeseen chain of events can start a fire. Even the best added-on safety technology can fail or be destroyed in the course of an accident.

An additional argument for inherent safety is its usefulness in meeting security threats. Add-on safety measures can often easily be deactivated by those who want to do so. When terrorists enter a chemical plant with the intent to blow it up, it does not matter much that all ignition sources have been removed from the vicinity of explosive materials (although this may perhaps have solved the safety problem). The perpetrators will bring their own ignition source. In contrast, most measures that make a plant inherently safer will also contribute to diverting terrorist threats. If the explosive substance has been replaced by a nonexplosive one or the inventories of explosive and inflammable substances have been drastically reduced, then the plant will be much less attractive to terrorists and therefore also a less likely target of attack (Hansson 2010a).

Most of the development of techniques for inherent safety has taken place within the chemical industry. Another major industry where inherent safety is often discussed is the nuclear industry, where it is referred to in efforts to construct new, safer types of reactors. A reactor will be inherently safer than those currently in use if, even in the case

of failure of all active cooling systems and complete loss of coolant, the temperatures will not be high enough to trigger the release of radioactive fission products (Brinkmann et al. 2006).

## Safety Factors

Probably, humans have made use of safety reserves since the origin of our species. We have added extra strength to our houses, tools, and other constructions in order to be on the safe side. The use of safety factors, i.e., numerical factors for dimensioning safety reserves, originated in the latter half of the nineteenth century (Randall 1976). Their use is now well established in structural mechanics and in its many applications in different engineering disciplines. Elaborate systems of safety factors have been specified in norms and standards (Clausen et al. 2006).

A safety factor is most commonly expressed as the ratio between a measure of the maximal load not leading to the specified type of failure and a corresponding measure of the maximal load that is expected to be applied. In some cases, it may instead be expressed as the ratio between the estimated design life and the actual service life. A safety factor is typically intended to protect against a specific integrity-threatening mechanism, and different safety factors can be used against different such mechanisms. Hence, one safety factor may be required for resistance to plastic deformation and another for fatigue resistance.

According to standard accounts of structural mechanics, safety factors are intended to compensate for five major categories of sources of failure:

1. Higher loads than those foreseen.
2. Worse properties of the material than foreseen.
3. Imperfect theory of the failure mechanism in question.
4. Possibly unknown failure mechanisms.
5. Human error (e.g., in design) (Knoll 1976; Moses 1997).

The first two of these can in general be classified as variabilities, that is, they refer to the variability of empirical indicators of the propensity for failure. They are therefore accessible to probabilistic assessment (although these assessments may be more or less uncertain). The last three failure types refer to eventualities that are difficult or impossible to represent in probabilistic terms and, therefore, belong to the category of (non-probabilizable) uncertainty. They are not easily amenable to probabilistic treatment. It is, for instance, difficult to see how a calculation could be accurately adjusted to compensate self-referentially for an estimated probability that it is itself wrong. However, these difficulties do not make these sources of failure less important. Safety factors are used to deal both with those failures that can be accounted for in probabilistic terms and those that cannot (Doorn and Hansson 2011).

## Multiple Independent Safety Barriers

Safety barriers are arranged in chains. The aim is to make each barrier independent of its predecessors so that if the first fails, then the second is still intact, etc. Typically, the first barriers are measures to prevent an accident, after which follow barriers that limit the consequences of an accident, and, finally, rescue services as the last resort.

The archetype of multiple safety barriers is an ancient fortress. If the enemy manages to pass the first wall, there are additional layers that protect the defending forces. Some engineering safety barriers follow the same principle of concentric physical barriers. Interesting examples of this can be found in nuclear waste management. The waste can for instance be put in a copper canister that is constructed to resist the foreseeable challenges. The canister is surrounded by a layer of bentonite clay that protects it against small movements in the rock and absorbs leaking radionuclides. This whole construction is placed in deep rock, in a geological formation that has been selected to minimize transportation to the surface of any possible leakage of radionuclides. The whole system of barriers is constructed to have a high degree of redundancy so that if one of the barriers fails the remaining ones will suffice. With the usual standards of probabilistic risk analysis, the whole series of barriers around the waste would not be necessary. Nevertheless, sensible reasons can be given for this approach, namely reasons that refer to uncertainty. Perhaps the copper canister will fail for some unknown reason not included in the calculations. Then, hopefully, the radionuclides will stay in the bentonite, etc.

The notion of multiple safety barriers can also refer to safety barriers that are not placed in a spatial sequence like the defense walls of a fortress but are arranged consecutively in a functional sense. The essential feature is that the second barrier is put to work when the first one fails, etc. Consider, for instance, the protection of workers against a dangerous gas such as hydrogen sulfide that can leak from a chemical process. An adequate protection against this danger can be constructed as a series of barriers. The first barrier consists in constructing the whole plant in a way that excludes uncontrolled leakage as far as possible. The second barrier is careful maintenance, including regular checking of vulnerable details such as valves. The third barrier is a warning system combined with routines for evacuation of the premises in the case of a leakage. The fourth barrier is efficient and well-trained rescue services.

The basic idea behind multiple barriers is that even if the first barrier is well constructed, it may fail, perhaps for some unforeseen reason, and that the second barrier should then provide protection. For a further illustration of this principle, suppose that a shipbuilder comes up with a convincing plan for an unsinkable boat. Calculations show that the probability of the ship sinking is incredibly low and that the expected cost per life saved by the lifeboats is above 1,000 million dollars, a sum that can evidently be more efficiently used to save lives elsewhere.

How should the naval engineer respond to this proposal? Should she accept the verdict of the probability calculations and the economic analysis, and exclude lifeboats from the design? There are good reasons why a responsible engineer should not act in this way: The calculations may possibly be wrong, and if they are, then the outcome may be disastrous. Therefore, the additional safety barrier in the form of lifeboats (and evacuation routines and all the rest) should not be excluded. Although the calculations indicate that such measures are inefficient, these calculations are not certain enough to justify such a decision. (This is a lesson that we should have learned from the *Titanic* disaster.)

The major problem in the construction of safety barriers is how to make them as independent of each other as possible. If two or more barriers are sensitive to the same type of impact, then one and the same destructive force can get rid of all of them in one swoop. Hence, three consecutive safety valves on the same tube may all be destroyed in a fire or they may all be incapacitated due to the same mistake by the maintenance department. It is essential, when constructing a system of safety barriers, to make the barriers as independent as possible. Often, more safety is obtained with fewer but independent barriers than with many that are sensitive to the same sources of incapacitation.

These three principles of engineering safety – inherent safety, safety factors, and multiple barriers – are quite different in nature, but they have one important trait in common: They all aim at protecting us not only against risks that can be assigned meaningful probability estimates, but also against dangers that cannot be probabilized, such as the possibility that some unforeseen event triggers a hazard that is seemingly under control. It remains, however, to investigate more in detail the principles underlying safety engineering and, not least, to clarify how they relate to other principles of engineering design.

## Ethics

---

Moral theorizing has mostly referred to the values of certain outcomes. The evaluation of uncertain outcomes is conventionally referred to decision theory, where it is treated as means-ends (instrumental) reasoning directed toward the attainment of given ends. Hence, moral philosophy refers primarily to human behavior in situations when the outcomes of actions are well defined and knowable. Decision theory takes assessments of these cases for given and derives from them assessments for situations involving risk and uncertainty. In this derivation, it operates, or so it is assumed, exclusively with criteria of rationality and does not add any new moral values. The dominating framework for these deliberations is expected utility theory.

Consider a person who risks a sleeping person's life by playing Russian roulette on her. In a moral assessment of this act, we need to consider (1) the set of consequences that will ensue if the person is killed, and (2) the set of consequences that will fall out if the person is not killed. In addition to this, we should also take into account (3) the act of risk imposition, which in this case takes the form of intentionally performing an act that may develop into an instance of either (1) or (2). In many people's moral appraisal of this misdeed, (3) has considerable weight. The act of deliberate risk-taking is perceived as a wrongdoing against the sleeping person, even if she is not killed and even if she never becomes aware of this episode or any disadvantage emanating from it. However, in the standard decision-theoretic approach, only (1) and (2) are taken into account (weighed according to their probabilities), whereas (3) is left out from the analysis.

This can be expressed with somewhat more precision in the following terminology: In a conventional decision-theoretical appraisal of this example, (1) and (2) will be replaced by their *closest deterministic analogs*. (1) is then evaluated as the act of discharging a fully loaded pistol at the sleeping person's head and (2) as that of letting off an unloaded pistol at her head. The composite act of performing what may turn out to be either (1) or (2) is assumed to have no other morally relevant aspects than those that are present in at least one of these two acts, both of which have well-determined consequences. The additional moral issues in (3), i.e., the issues concerning risk-taking per se, have no place in this account.

It is a general feature of this form of decision-theoretical analysis that if an act has moral aspects that are not present in the closest deterministic analog of any of its alternative developments, then these aspects are left out from the analysis. The crucial (but usually unstated) underlying assumption is that an adequate appraisal of an action under risk or uncertainty can be based on the values that pertain to its closest deterministic analogs. But as we saw from the Russian roulette example, this assumption has the disadvantage of excluding from our consideration the moral implications of risk-taking per se. This exclusion is unavoidable since risk-taking is by definition absent from the closest deterministic analogs that are used in the analysis.

The exclusion of risk-taking from consideration in most of moral theory can be clearly seen from the deterministic assumptions commonly made in the standard type of life-or-death examples that are used to explore the implications of moral theories. In the famous trolley problem, you are assumed to know that if you flip the switch, then one person will be killed, whereas if you do not flip it, then five other persons will be killed (Foot 1967). In Thomson's (1971) "violinist" thought experiment, you know for sure that the violinist's life will be saved if he is physically connected to you for 9 months, otherwise not. In Williams's (1973) example of Jim and the Indians, Jim knows for sure that if he kills 1 Indian, then the commander will spare the lives of 19 whom he would otherwise kill, etc. This is in stark contrast to ethical quandaries in real life, where action problems with human lives at stake seldom come with certain knowledge of the consequences of the alternative courses of action. Instead, uncertainty about the consequences of one's actions is a major complicating factor in most real-life dilemmas.

There are no easy answers to questions such as what risks you are allowed to impose on one person in order to save another or what risks a person can be morally required to take in order to save a stranger. These are questions that present themselves to us as moral questions, not as issues for decision-theoretical reckoning to take place after the moral deliberations have been finished. The exclusion of such issues from most discussions in moral philosophy has the effect of removing essential aspects of actual moral decision-making from our deliberations on moral theory. In order to include them, we have to give up the traditional assumption that the valuation of risk should take the form of applying decision-theoretical – and thus nonmoral – reasoning to values that refer to the moral evaluation of non-risky outcomes (in the form of closest deterministic analogs). Instead, we have to treat risk-taking per se as an object of moral appraisal.

The obvious way to develop an ethical theory of risk would be to generalize one of the existing ethical theories so that it can be effectively applied to situations involving risk. The problem of how to perform this generalization can be specified in terms of *the causal dilution problem*. It was presented by Robert Nozick (1974) as a problem for deontological ethics but is equally problematic for other moral theories.

► *The causal dilution problem (general version):*

Given the moral appraisals that a moral theory  $T$  makes of value-carriers with well-determined properties, what moral appraisals does (a generalized version of)  $T$  make of value-carriers whose properties are not well-determined beforehand?

In utilitarian moral theory, one fairly obvious approach to the causal dilution problem for utilitarianism is the following (Carlson 1995):

► *Actualism*

The utility of a (probabilistic) mixture of potential outcomes is equal to the utility of the outcome that actually materializes.

To exemplify the actualist approach, consider an engineer's decision whether or not to reinforce a bridge before it is being used for a single, very heavy transport. There is a 50% risk that the bridge will collapse if it is not reinforced. Suppose that she decides not to reinforce the bridge and that everything goes well; the bridge is not damaged. According to the actualist approach, what she did was right. This is, of course, in stark contrast to common moral intuitions.

But actualism is not the standard decision-theoretical solution to the causal dilution problem for utilitarianism. The standard approach is to maximize expected utility:

► *Expected utility:*

The utility of a probabilistic mixture of potential outcomes is equal to the probability-weighted average of the utilities of these outcomes.

This is a much more credible solution, and it has the important advantage of being a fairly safe method to maximize the outcome in the long run. Suppose, for instance, that the expected number of deaths in traffic accidents in a region will be 300 per year if safety belts are compulsory and 400 per year if they are optional. Then, if these calculations are correct, about 100 more persons per year will actually be killed in the latter case than in the former. We know, when choosing one of these options, whether it will lead to fewer or more deaths than the other option. If we aim at reducing the number of traffic casualties, then this can, due to the law of large numbers, safely be achieved by maximizing the expected utility (i.e., minimizing the expected number of deaths).

However, this argument is not valid for case-by-case decisions on unique or very rare events. Suppose, for instance, that we have a choice between a probability of 0.001 of an event that will kill 50 persons and a 0.1 probability of an event that will kill one person. Here, random effects will not be leveled out as in the safety belt case. In other words, we do not know, when choosing one of the options, whether or not it will lead to fewer deaths than the other option. In such a case, taken in isolation, there is no compelling reason to maximize expected utility.

Even when the leveling-out argument for expected utility maximization is valid, compliance with this principle is not required by rationality. It is quite possible for a rational agent to refrain from minimizing total damage in order to avoid imposing high-probability risks on individuals. This can be exemplified with an example involving an acute situation in a chemicals factory (Hansson 1993). There are two ways to repair a serious gas leakage that threatens to develop into a disaster. One of the options is to send in the repairman immediately. (There is only one person at hand who is competent to do the job.) He will then run a risk of 0.9 to die due to an explosion of the gas immediately after he has performed the necessary technical operations. The other option is to immediately let out gas into the environment. In that case, the repairman will run no particular risk, but each of 10,000 persons in the immediate vicinity of the plant runs a risk of 0.001 to be killed by the toxic effects of the gas. The maxim of maximizing expected utility requires that we send in the repairman to die. This is also a fairly safe way to minimize the number of actual deaths. However, it is not clear that it is the only possible response that is rational. A rational decision-maker may refrain from maximizing expected utility (minimizing expected damage) in order to avoid what would be unfair to a single individual and infringe her rights. Hence, we have to go beyond expected utility theory in order to do justice to important moral intuitions about the rights of individuals.

As already mentioned, the causal dilution problem was originally formulated for rights-based theories by Robert Nozick. He asked: "Imposing how slight a probability of a harm that violates someone's rights also violates his rights?" (Nozick 1974, p. 7). In somewhat more general language, we can restate the question as follows:

► *The causal dilution problem for deontological/rights-based moral theories:*

Given the duties/rights that a moral theory *T* assigns with respect to actions with well-determined properties, what duties/rights does (a generalized version of) *T* assign with respect to actions whose properties are not well-determined beforehand?

A rights-based moral theory can be extended to indeterministic cases by just prescribing that if A has a right that B does not bring about a certain outcome, then A also has a right that B does not perform any action that has a nonzero risk of bringing about that outcome. Unfortunately, such a strict extension of rights and prohibitions is socially untenable. Your right not to be killed by me certainly implies a prohibition for me to perform certain acts that involve a risk of killing you, but it cannot prohibit all such acts. Such a strict interpretation would make human society impossible. For instance, you would not be allowed to drive a car in the town where I live since this increases my risk of being killed by you.

Hence, rights and prohibitions have to be defeasible so that they can be canceled when probabilities are small. The most obvious way to achieve this is to assign to each right (prohibition) a probability limit. Below that limit, the right (prohibition) is canceled. However, as Nozick observed, such a solution is not credible since probability limits “cannot be utilized by a tradition which holds that stealing a penny or a pin or anything from someone violates his rights. That tradition does *not* select a threshold measure of harm as a lower limit, in the case of harms certain to occur” (Nozick 1974, p. 75).

Clearly, a moral theory need not treat a slight probability of a sizable harm in the same way that it treats a slight harm. The analogy is nevertheless relevant. The same basic property of traditional rights theories, namely the uncompromising way in which they protect against disadvantages for one person inflicted by another, prevents them from drawing a principled line either between harms or between probabilities in terms of their acceptability or negligibility. In particular, since no rights-based method for the determination of such probability limits seems to be available, they would have to be external to the rights-based theory. Exactly the same problem obtains for deontological theories.

Finally, let us consider contract theories. They may perhaps appear somewhat more promising. The criterion that they offer for the deterministic case, namely consent among all those involved, can also be applied to risky options. Can we then solve the causal dilution problem for contract theories by saying that risk impositions should be accepted to the degree that they are supported by a consensus?

Unfortunately, this solution is fraught with problems. Consent, as conceived in contract theories, is either actual or hypothetical. Actual consent does not seem to be a realistic criterion in a complex society in which everyone performs actions with marginal but additive effects on many other people's lives. According to the criterion of actual consent, you have a veto against me or anyone else who wants to drive a car in the town where you live. Similarly, I have a veto against your use of (any type of) fuel to heat your house since the emissions contribute to health risks that affect me. In this way, we can all block each other, creating a society of stalemates. When all options in a decision are associated with risks, and all parties claim their rights to keep clear of the risks that others want to impose on them, the criterion of actual consent does not seem to be of much help.

We are left then with hypothetical consent. However, as the debate following John Rawls's *Theory of Justice* has shown, there is no single decision rule for risk and uncertainty that all participants in a hypothetical initial situation can be supposed to adhere to (Hare 1973; Harsanyi 1975). It remains to show that a viable consensus on risk impositions can be reached among participants who apply different decision rules in situations of risk and uncertainty (If a unanimous decision is reached due to the fact that everybody applies the same decision rule, then the problem has not been solved primarily by contract theory but by the underlying theory for individual decision-making). Apparently, this has not been done and, hence, contract theory does not either have a solution to the causal dilution problem.

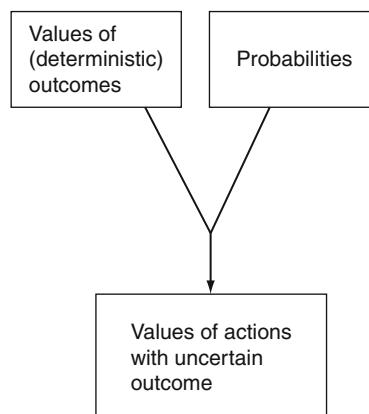
The difficulties that we encounter when trying to solve the causal dilution problem within the frameworks of the common types of moral theories are indications of a deeper problem. The attempted solutions reviewed above are all based on an implicit derivation principle: It is assumed that if the moral appraisals of actions with deterministic outcomes are given, then we can derive from them moral appraisals of actions whose outcomes are probabilistic mixtures of such deterministic outcomes. In utilitarian approaches, it is furthermore assumed that probabilities and (deterministic) utilities are all the information that we need (☞ Fig. 2.3). However, this picture is much too simplified. The morally relevant aspects of situations of risk and uncertainty go far beyond the impersonal, free-floating sets of consequences that decision theory operates on. Risks are inextricably connected with interpersonal relationships. They do not just “exist”; they are taken, run, or imposed (Thomson 1985a). To take just one example, it makes a moral difference if it is one’s own life or that of somebody else that one risks in order to earn a fortune for oneself. Therefore, person-related aspects such as agency, intentionality, consent, etc., will have to be taken seriously in any reasonably accurate account of real-life indeterminism (☞ Fig. 2.4).

Based on this analysis, the causal dilution problem can be replaced by a *defeasance problem* that better reflects the moral issues of risk impositions:

► *The defeasance problem:*

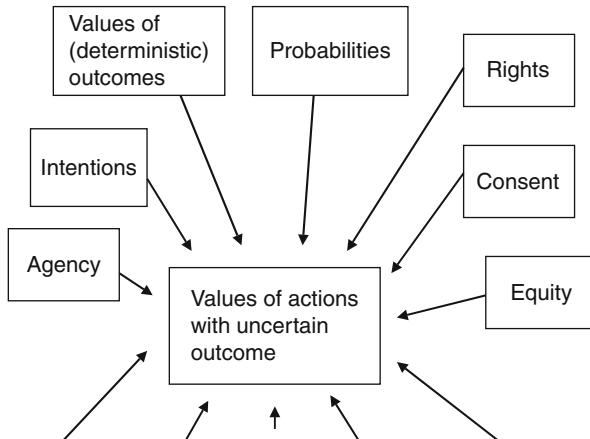
It is a *prima facie* moral right not to be exposed to risk of negative impact, such as damage to one’s health or one’s property, through the actions of others. What are the conditions under which this right is defeated so that someone is allowed to expose other persons to risk?

The defeasance problem is a truly moral problem, not a decision-theoretical one. As far as I can see, it is the central ethical issue that a moral theory of risk has to deal with. Obviously, there are many ways to approach it, only few of which have been developed. It remains to investigate and compare the various solutions that are possible. My own preliminary solution refers to reciprocal exchanges of risks and benefits. Each of us takes risks in order to obtain benefits for ourselves. It is beneficial for all of us to extend this practice to mutual exchanges of risks and benefits. If others are allowed to drive a car, exposing you to certain risks, then in exchange you



■ Fig. 2.3

The traditional account of how values of indeterministic outcomes can be derived

**Fig. 2.4**

A less incomplete picture of the influences on the values of indeterministic options

are allowed to drive a car and expose them to the corresponding risks. This (we may suppose) is to the benefit of all of us. In order to deal with the complexities of modern society, we also need to apply this principle to exchanges of different types of risks and benefits. We can then *regard exposure of a person to a risk as acceptable if it is part of a social system of risk-taking that works to her advantage and gives her a fair share of its advantages*.

This solution is only schematic, and it gives rise to further problems that need to be solved. Perhaps the most difficult of these problems is how to deal with large differences among the members of society in their assessments of risks and benefits. But with the approach presented here, we have, or at least so I wish to argue, a necessary prerequisite in place, namely, the right agenda for the ethics of risk. According to traditional risk analysis, in order to show that it is acceptable to impose a risk on Ms. Smith, the risk-imposer only has to give sufficient reasons for accepting the risk as such, as an impersonal entity. According to the proposal just presented, this is not enough. The risk-imposer has to give sufficient reasons why Ms. Smith – as the particular person that she is – should be exposed to the risk. This can credibly be done only by showing that this risk exposure is part of some arrangement that works to her own advantage. For a more detailed discussion of this approach, see Hansson (2003b).

## Philosophy of Economics

Risks have a central role in economic theory, and there are obvious parallels between the problems of economic risk and the problems concerning other types of risk such as risks to health and the environment. Let us have a look at two interesting issues in the philosophy of economic risk: the aggregation problem and the problem of positive risk-taking.

*The aggregation problem* concerns how we compare risks accruing to different individuals. Standard risk analysis follows the principles of classical utilitarianism. All risks are summed up in one and the same balance irrespectively of whom they accrue to. Thus, all risks are taken to be fully comparable and additively aggregable. In risk-benefit analysis, benefits are added in the

same way and, finally, the sum of benefits is compared to the sum of risks in order to determine whether the total effect is positive or negative. In such a model, just as in classical utilitarianism, individuals have no other role than as carriers of utilities and disutilities, the values of which are independent of whom they are carried by.

An obvious alternative to this utilitarian approach is to treat each individual as a separate moral unit. Then risks and benefits pertaining to one and the same person can be weighed against each other, whereas risks and benefits for different persons are added or otherwise aggregated since they are considered to be incomparable. Such “individualistic” risk weighing is quite different from the total aggregations that are standard in risk analysis. But individualistic risk weighing dominates in medicine. It is applied for instance in ethical evaluations of clinical trials. It is an almost universally accepted principle in research ethics that a patient should not be included in a clinical trial unless there is genuine uncertainty on whether or not participation in the trial is better for her than the standard treatment that she would otherwise receive. That her participation is beneficial for others (such as future patients) cannot outweigh a negative net effect on her own health; in other words, her participation has to be supported by an appraisal that is restricted to risks and benefits for herself (London 2001; Hansson 2004a).

The two traditions in risk assessment differ in the same way as the “old” and “new” schools of welfare economics. In Arthur Pigou’s so-called old welfare economics, the values pertaining to different individuals are added up to one grand total. This is also the approach of mainstream risk analysis. The new school in welfare economics that dominates mainstream economics since the 1930s refrains from adding individual values. Instead, it treats the welfare of different individuals as incomparable. This became the standard approach after Lionel Robbins had shown how economic analysis can dispense with interpersonal comparability (Pareto optimality is the central tool needed to achieve this). The individualist approach that was exemplified above with clinical trials is based on the same basic principles as those applied by the new school in welfare economics (Hansson 2006a).

Mainstream risk analysis and mainstream economics represent two extremes with respect to interindividual comparisons. The aggregations of total risk that are performed routinely in risk analysis stand in stark contrast to the consistent avoidance of interindividual comparisons that is a major guiding principle in modern economics. This difference also has repercussions in the ideological uses of the respective disciplines. It is an implicit message of risk-benefit analysis that a rational person should accept being exposed to a risk if this brings greater benefits for others. The implicit message of modern (new school) welfare economics is much more appreciative of self-interested behavior.

The issue of *positive risk-taking* appears to be more or less specific for economic risks. Risk is by definition undesirable, and we expect a rational person to avoid risk as far as possible. But in economics, risk-taking is often considered to be desirable. The capitalist’s risk taking is acknowledged as essential for the efficiency of a capitalist system, and it is also taken to justify the owner’s prerogative to exert the ultimate control over companies and to reap the profits. As said already by Adam Smith in his *Wealth of Nations*, “something must be given for the profits of the undertaker of the work who hazards his stock in this adventure” (Smith [1776] 1976, p. 1:66).

The risk taking that Smith referred to was a substantial one, namely the risk of bankruptcy. According to Smith, becoming bankrupt is “perhaps the greatest and most humiliating calamity which can befall an innocent man” (Smith [1776] 1976, p. 1:342). This was the risk that the capitalist was supposed to take and to be compensated for. Its seriousness was essential for Smith’s argument as we can see from his negative attitude to arrangements that reduce the

risk from that of bankruptcy to that of losing the invested capital. The most important such arrangement was the joint-stock company (with limited liability) to which Smith was decidedly averse (Smith [1776] 1976, p. 2:741).

But since Smith's time capitalism has been fundamentally transformed, two major reductions in capitalist risk-taking have taken place. The first of these occurred in the latter half of the nineteenth century when corporations with limited liability became the dominant legal form of private companies in the industrialized parts of the world (Handlin and Handlin 1945; Prasch 2004). Due to the massive spread of limited liability, personal risk taking in most major industrial and financial endeavors was brought down from bankruptcy to loss of the original investment, which was exactly what Adam Smith had warned against.

The second reduction in economic risk-taking took place about 100 years later. Beginning in the late twentieth century, private investment in companies has to an increasing extent been mediated by institutions and funds that diversify their securities in a sophisticated way to reduce risk taking. Portfolio theory and modern financial marketplaces have combined to make risk spreading much more efficient than what was previously possible. Today, an owner who has applied prudent risk spreading only runs a risk that approximates the general background risk of the economy. In terms of risk-taking, his situation is arguably less akin to that of businesspeople risking everything they own than to that of the nineteenth century landlords who according to John Stuart Mill "grow richer, as it were in their sleep, without working, risking, or economizing" (Mill [1848] 1965, pp. 3:819–820).

Risk-spread ownership has thoroughly transformed the economic system, but its philosophical implications have not been much discussed. It is for instance not unreasonable to ask what effects this development has on the legitimacy of the owner's prerogative that was previously based at least in part on the risk-taking role of owner.

## Political Philosophy

---

Although risk and uncertainty are ubiquitous in political and social decision-making, there has been very little contact between risk studies and more general studies of social decision processes. Public discussions of risk are dominated by a way of thinking that is markedly different from how democratic decision-making is commonly discussed. We can see this from the frequent references in discussions on risk to three terms that are not much used in general discussions on democratic decision-making.

"Consent" is one of these words. Consent by the public is often taken to be the goal of public communications on risk. The following quotation is not untypical:

- ▶ Community groups have in recent years successfully used zoning and other local regulations, as well as physical opposition (e.g., in the form of sitdowns or sabotage), to stall or defeat locally unacceptable land uses. In the face of such resistance, it is desirable (and sometimes even necessary) to draw forth the consent of such groups to proposed land uses (Simmons 1987, p. 6).

To consent means in this context "voluntarily to accede to or acquiesce in what another proposes or desires" (Oxford English Dictionary). This is very different from the role of the citizen as a decision-maker in a democracy.

The second of these words is "acceptance." The goal of risk communication is often taken to be public acceptance of a presumably rational act of risk-taking. Hence, in discussions on the

siting of potentially dangerous industries, “public acceptance” is usually taken to be the crucial criterion. This usage signals the same limitation in public participation as the word “consent.”

The third word is “trust.” Much of the discussion in the risk-related academic literature on the relationship between decision-makers and the public takes the public’s trust in decision-makers to be the obvious criterion of a well-functioning relationship. This is, again, very different from discussions on democracy in political philosophy. In a democratic constitution, the aim is the public’s democratic control over decision-makers, rather than their trust in them (Hayenjelm 2007).

In contrast to the limited approach to public participation in risk decisions that is indicated by the keywords “consent,” “acceptance,” and “trust,” let us consider the ideal of full democratic participation in decision-making. This ideal was very well expressed by Condorcet in his vindication of the French constitution of 1793. Condorcet divided decision processes into three stages. In the first stage, one “discusses the principles that will serve as the basis for decision in a general issue; one examines the various aspects of this issue and the consequences of different ways to make the decision.” At this stage, the opinions are personal, and no attempts are made to form a majority. After this follows a second discussion in which “the question is clarified, opinions approach and combine with each other to a small number of more general opinions.” In this way, the decision is reduced to a choice between a manageable set of alternatives. The third stage consists of the actual choice between these alternatives (Condorcet [1793] 1847, pp. 342–343).

The discussion on public participation in issues of risk has mostly been restricted to Condorcet’s third stage, and – as we have seen – often also to a merely confirming or consenting role in that stage. This approach is obviously untenable if we wish to see decisions on risk in the full context of public decision-making in a democratic society. Without public participation at all three stages of the decision-making process, risk issues cannot be dealt with democratically. Therefore, the discussion needs to be shifted away from special procedures for dealing with risk. Instead our focus should be on how the special characteristics of risk-related issues can best be dealt with in our general decision-making processes.

## Further Research

Far from being an unusual oddity in philosophy, the topic of risk connects directly to central issues in quite a few subdisciplines of philosophy. In some of these subdisciplines, a significant amount of research on risk has already taken place. In others, only the first steps toward systematic studies of risk have been taken. But in all of them, important philosophical issues related to risk remain unexplored. The following are ten of the most important issues for further research that have been pointed out above:

- When is trust in experts on risks justified, and when is distrust irrational?
- How remote possibilities of disaster should be taken seriously?
- How can we account for probabilistic reasoning that seems rational but is not compatible with the standard theory of probability?
- To what extent, and in what ways, should practical consequences have influence on scientific assessments of risk?
- How can the principles of safety engineering be accounted for, and how do they relate to probabilistic risk analysis?

- How can a risk or safety analysis take into account the possibility that the analysis itself is wrong?
- When is it ethically permissible to expose another person to a risk?
- How can utilitarianism be extended or adjusted so that it provides us with a reasonable account of the ethics of risk-taking?
- Do we have a right not to be exposed to risks and, in that case, when can it be overruled?
- What role should those exposed to a risk have in democratic decisions on that risk?

## References

---

- Ahteensuu M (2008) In dubio pro natura? PhD thesis in philosophy, University of Turku
- Blackburn S (1973) Reason and prediction. Cambridge University Press, Cambridge
- Brinkmann G, Pirson J, Ehster S, Dominguez MT, Mansani L, Coe I, Moormann R, Van der Mheen W (2006) Important viewpoints proposed for a safety approach of HTGR reactors in Europe. Final results of the EC-funded HTR-L project. *Nucl Eng Des* 236:463–474
- Burgos R, Defeo O (2004) Long-term population structure, mortality and modeling of a tropical multi-fleet fishery: the red grouper Epinephelus morio of the Campeche bank, Gulf of Mexico. *Fish Res* 66:325–335
- Carlson E (1995) Consequentialism reconsidered. Kluwer, Dordrecht/Boston
- Clausen J, Hansson SO, Nilsson F (2006) Generalizing the safety factor approach. *Reliab Eng Syst Saf* 91:964–973
- Cohen BL (2003) Probabilistic risk analysis for a high-level radioactive waste repository. *Risk Anal* 23:909–915
- Condorcet ([1793] 1847) Plan de Constitution, présenté à la convention nationale les 15 et 16 février 1793. *Oeuvres* 12:333–415
- Cox R, Winkler R (2010) Spill may prompt energy mergers. *New York Times* June 2, 2010. <http://www.nytimes.com/2010/06/03/business/03views.html>. Accessed 9 June 2011
- Cranor CF (1997) The normative nature of risk assessment: features and possibilities. *Risk Health Saf Environ* 8:123–136
- Cranor CF, Nutting K (1990) Scientific and legal standards of statistical evidence in toxic tort and discrimination suits. *Law Philos* 9:115–156
- Donahoe FJ (1969) 'Anomalous' water. *Nature* 224:198
- Doorn N, Hansson SO (2011) Should safety factors replace probabilistic design? *Philos Technol* 24:151–168
- Feeleppa R (1981) Epistemic utility and theory acceptance: comments on Hempel. *Synthese* 46:413–420
- Fischhoff B (1977) Perceived informativeness of facts. *Hum Percept Perform* 3(2):349–358
- Fischhoff B, Lichtenstein S, Slovic P, Derby SL, Keeney RL (1981) Acceptable risk. Cambridge University Press, Cambridge
- Foot P (1967) The problem of abortion and the doctrine of the double effect. *Oxford Rev* 5:5–15. Reprinted in her *Virtues and Vices*, Oxford: Basil Blackwell, 1978
- Handlin O, Handlin MF (1945) Origins of the American business corporation. *J Econ Hist* 5:1–23
- Hansson SO (1993) The false promises of risk analysis. *Ratio* 6:16–26
- Hansson SO (1995) The detection level. *Regul Toxicol Pharmacol* 22:103–109
- Hansson SO (1996) Decision-making under great uncertainty. *Philos Soc Sci* 26:369–386
- Hansson SO (1998) Setting the limit: occupational health standards and the limits of science. Oxford University Press, New York/Oxford
- Hansson SO (2002) Replacing the no effect level (NOEL) with bounded effect levels (OBEL and LEBEL). *Stat Med* 21:3071–3078
- Hansson SO (2003a) Are natural risks less dangerous than technological risks? *Philos Nat* 40:43–54
- Hansson SO (2003b) Ethical criteria of risk acceptance. *Erkenntnis* 59:291–309
- Hansson SO (2004a) Weighing risks and benefits. *Topoi* 23:145–152
- Hansson SO (2004b) Fallacies of risk. *J Risk Res* 7:353–360
- Hansson SO (2004c) Philosophical perspectives on risk. *Techne* 8(1):10–35
- Hansson SO (2004d) Great uncertainty about small things. *Techne* 8(2):26–35
- Hansson SO (2005) Seven myths of risk. *Risk Manage* 7(2):7–17
- Hansson SO (2006a) Economic (ir)rationality in risk analysis. *Econ Philos* 22:231–241
- Hansson SO (2006b) How to define – a tutorial. *Princípios, Revista de Filosofia* 13(19–20):5–30

- Hansson SO (2007a) Philosophical problems in cost-benefit analysis. *Econ Philos* 23:163–183
- Hansson SO (2007b) Values in pure and applied science. *Found Sci* 12:257–268
- Hansson SO (2008) Regulating BFRs – from science to policy. *Chemosphere* 73:144–147
- Hansson SO (2009a) Should we protect the most sensitive people? *J Radiat Prot* 29:211–218
- Hansson SO (2009b) Measuring uncertainty. *Studia Log* 93:21–40
- Hansson SO (2010a) Promoting inherent safety. *Process Saf Environ Prot* 88:168–172
- Hansson SO (2010b) Past probabilities. *Notre Dame J Formal Logic* 51:207–233
- Hansson SO, Rudén C (2006) Evaluating the risk decision process. *Toxicology* 218:100–111
- Hare RM (1973) Rawls's theory of justice. *Am Philos Quart* 23:144–155 and 241–252
- Harsanyi JC (1975) Can the maximin principle serve as a basis for morality – critique of Rawls, J theory. *Am Pol Sci Rev* 69(2):594–606
- Harsanyi JC (1983) Bayesian decision theory, subjective and objective probabilities, and acceptance of empirical hypotheses. *Synthese* 57:341–365
- Hayenjelm M (2007) Trusting and taking risks: a philosophical inquiry. Ph.D. thesis, KTH, Stockholm
- Hempel CG (1960) Inductive inconsistencies. *Synthese* 12:439–469
- International Organization for Standardization (2002) Risk management – vocabulary – guidelines for use in standards, ISO/IEC Guide 73/2002
- Knoll F (1976) Commentary on the basic philosophy and recent development of safety margins. *Can J Civ Eng* 3:409–416
- Krewski D, Goddard MJ, Murdoch D (1989) Statistical considerations in the interpretation of negative carcinogenicity data. *Regul Toxicol Pharmacol* 9:5–22
- Leisenring W, Ryan L (1992) Statistical properties of the NOAEL. *Regul Toxicol Pharmacol* 15:161–171
- Levi I (1962) On the seriousness of mistakes. *Philos Sci* 29:47–65
- Levi I (1973) Gambling with truth. MIT Press, Cambridge, MA
- London AJ (2001) Equipoise and international human-subjects research. *Bioethics* 15:312–332
- Lopez RE, Holle RL (1998) Changes in the number of lightning deaths in the United States during the twentieth century. *J Climate* 11:2070–2077
- MacLean D (ed) (1985) Values at risk. Rowman & Allanheld, Totowa
- Mill JS ([1848] 1965) The principles of political economy with some of their applications to social philosophy. In: Robson JM (ed) Collected works of John Stuart Mill, vol 2–3. University of Toronto Press, Toronto
- Miller CO (1988) System safety. In: Wiener EL, Nagel DC (eds) Human factors in aviation. Academic, San Diego, pp 53–80
- Möller N, Hansson SO (2008) Principles of engineering safety: risk and uncertainty reduction. *Reliab Eng Syst Saf* 93:776–783
- Möller N, Hansson SO, Peterson M (2006) Safety is more than the antonym of risk. *J Appl Philos* 23(4):419–432
- Moses F (1997) Problems and prospects of reliability-based optimisation. *Eng Struct* 19:293–301
- National Research Council (1983) Risk assessment in the federal government: managing the process. National Academy Press, Washington, DC
- Nozick R (1974) Anarchy, state, and utopia. Basic Books, New York
- O'Riordan T, Cameron J (eds) (1994) Interpreting the precautionary principle. Earthscan, London
- O'Riordan T, Cameron J, Jordan A (eds) (2001) Reinterpreting the precautionary principle. Cameron May, London
- Prasch RE (2004) Shifting risk: the divorce of risk from reward in American capitalism. *J Econ Issues* 38:405–412
- Rabinowicz W (2002) Does practical deliberation crowd out self-prediction? *Erkenntnis* 57:91–122
- Randall FA (1976) The safety factor of structures in history. *Prof Saf* 1976(January):12–28
- Roeser S (2006) The role of emotions in judging the moral acceptability of risks. *Saf Sci* 44:689–700
- Royal Society (1983) Risk assessment. Report of a Royal Society Study Group, London
- Rudén C, Hansson SO (2008) Evidence based toxicology – ‘sound science’ in new disguise. *Int J Occup Environ Health* 14:299–306
- Sandin P (1999) Dimensions of the precautionary principle. *Hum Ecol Risk Assess* 5:889–907
- Shrader-Frechette K (1991) Risk and rationality: philosophical foundations for populist reforms. University of California Press, Berkeley
- Simmons J (1987) Consent and fairness in planning land use. *Bus Prof Ethics J* 6(2):5–20
- Smith A ([1776] 1976) An inquiry into the nature and causes of the wealth of nations. In: Campbell RH, Skinner AS, Todd WB (eds) The Glasgow edition of the works and correspondence of Adam Smith, vol 2. Clarendon, Oxford
- Spohn W (1977) Where Luce and Krantz do really generalize Savage's decision model. *Erkenntnis* 11:113–134
- Tench W (1985) Safety is no accident. Collins, London
- Thomson JJ (1971) A defense of abortion. *Philos Public Aff* 1:47–66
- Thomson JJ (1985a) Imposing risk. In: Gibson M (ed) To breathe freely. Rowman & Allanheld, Totowa, pp 124–140

- Thomson PB (1985b) Risking or being willing: Hamlet and the DC-10. *J Value Inquiry* 19:301–310
- Walton DN (1987) Informal fallacies: towards a theory of argument criticisms. *J. Benjamins, Amsterdam*
- Williams B (1973) A critique of utilitarianism. In: Smart JJC, Williams B (eds) *Utilitarianism: for and against*. Cambridge University Press, London

# 3 The Concepts of Risk and Safety

Niklas Möller

University of Cambridge, Cambridge, UK

<i>Introduction</i> .....	56
<i>Risk Perspectives</i> .....	57
<i>Notions of Risk</i> .....	58
Five Definitions of “Risk” .....	58
Acceptable Risk .....	60
Is Safety the Antonym of Risk? .....	60
<i>Aspects of Risk</i> .....	62
Harm and Probability .....	63
Risk as the Expected Value of Harm .....	64
Uncertainty .....	65
Further Aspects of Risk and Safety .....	67
<i>The Nature of Risk and Safety</i> .....	70
Risk and Normativity .....	71
Thick Concepts and Reductionism .....	74
Risk, Objectivity, and Social Constructions .....	77
<i>Further Research</i> .....	80
Analysis of Key Concepts .....	80
Further Issues .....	81
<i>Conclusion</i> .....	82

**Abstract:** The aim of this chapter is to analyze the concepts of risk and safety in the context of societal decision-making. Risk and safety research is a heterogeneous field, and different areas have conceived of the nature of risk in different ways. In the chapter, I categorize risk perspectives in three broad groups: the scientist approach, the psychological approach, and the cultural approach to risk. Between these groups, the nature and status of risk and safety have been the debated subjects. I will attempt to bring some light onto complicated and controversial philosophical topics such as whether risk and safety are natural or normative notions, whether they are social constructions, objective, or even real. This investigation will focus on a range of different questions. I will distinguish between five common definitions of the term “risk,” as well as contrast the notion of *risk* with both the notion of *safety* and the notion of *acceptable risk*. The main part of the chapter will focus on a quantitative or comparative concept of risk, that is, a notion that is in play in statements such as “the risk of flying is lower than the risk of traveling by car” and “the risk of nuclear power is  $10^{-4}$  deaths per reactor year.” The central aspects of such a notion of risk and safety will be discussed, in particular the notions of probability and harm. I will also discuss the common claim that it is the expectation value of the severity of harm that is the correct measure of risk. Furthermore, I investigate additional aspects such as epistemic uncertainty and other, more controversial aspects that have been proposed.

## Introduction

---

Risk research is a discipline in rapid development with contributors from many areas of the natural and social sciences. This reflects a growing concern about risks in society. Both professional and non-professional awareness of risks are increasing, and much effort is put into risk assessment, risk management, and risk communication. As a consequence, there are now well-developed societal practices in place involving risk and safety. Still, the central concepts of risk and safety remain somewhat unclear. When characterized, risk and safety are often treated either as relatively straightforward natural science concepts, or, to the contrary, as fundamentally subjective notions ill fitting for scientific study. However, without an in-depth understanding of its central concepts, the subject matter of risk and safety research remains fuzzy and it is unclear what the objective of reducing risk and achieving safety really amounts to.

The aim of this chapter is to analyze the concepts of risk and safety in the context of societal decision-making. This investigation will focus on a range of different questions. In the first section, I will categorize risk perspectives in three groups: the scientist approach, the psychological approach, and the cultural approach to risk. It aims to give an initial background and show the different points of departure that are characteristic for the heterogeneous field of risk research, and to supply an initial context for us to orient from and relate to in what follows.

In the second section, I distinguish between several meanings of the term “risk,” and discuss the notion of acceptable risk as well as the relation between the notions of risk and safety.

In the third section, I will discuss aspects of the quantitative concept of risk that will be in focus for the remainder of the chapter. I will investigate the fundamental aspects of probability and severity of harm, which are part of most quantitative conceptions of risk and safety. Furthermore, I will investigate additional aspects such as epistemic uncertainty and other, more controversial aspects that have been proposed.

The topic of the fourth section is the nature of risk and safety, by which is meant the status of the concepts: are risk and safety or normative notions, are they objective, or are they social constructions? Are they even real? The aim of the fifth section is to shed some light onto these complicated and controversial philosophical topics.

I suggest some topics for further research in the sixth section, before I end with a short conclusion in the final section.

## Risk Perspectives

---

Several different fields of investigation have been interested in the concept of risk: engineering, economics, political science, sociology, psychology, and philosophy, to name just a few. This heterogeneous research arena has resulted in many different perspectives on risk. Some theorists have grouped these perspectives rather finely. Ortwin Renn, for example, has divided the risk approaches into seven categories: (1) the actuarial approach, (2) the toxicological and epidemiological approach, (3) the engineering approach, (4) the economical approach, (5) the psychological approach, (6) social theories of risk, and (7) cultural theory of risk (Renn 1992, p. 56). (Even a fine-grained categorization such as this is incomplete. As indicated by the present volume, *philosophy of risk* is a growing area of research; cf. e.g., Shrader-Frechette 1993; Lewens 2007; Asveld and Roeser 2009; Hansson 1998, 2009). Often, however, theorists are satisfied with a less fine-grained grouping, depending on the task at hand.

For the purpose of this chapter, let us distinguish between three broad approaches to risk. The first perspective can be called the *scientist approach* to risk. The basic idea is that risk is a phenomenon that may be investigated like most phenomena in science, that is, by employing the scientific method. Risk is something that can, at least in principle, be measured in a systematic way, and the main task of the researcher is to find a sufficiently precise measure of the phenomena and to find ways of reducing the risk as much as possible. On most interpretations, the first four of Renn's perspectives would belong to this category. Statistical and probabilistic tools are important in this perspective as ways of measuring and describing the risks, in addition to investigations into the causal mechanisms of various risk-related phenomena that are at the core of both the toxicological-epidemiological and the engineering approaches (Renn 1992).

The second perspective may be labeled the *psychological approach* to risk. The basic interest in this perspective is to study people's *perceptions* of risk, that is, people's beliefs about risks and their way of relating to them. A dominant psychological method is psychometrical research in which the researcher tries to establish reliable measures of risk perceptions. The aim of the approach is to get a clear and distinct picture of how people estimate risks and how they make choices in relation to them – in particular, what influences whether they deem a risk acceptable or not. Various attitudes toward risk in general – especially risk-aversive and risk-seeking behavior – as well as what types of risk we deem more important than others, are typical topics of interest on this approach (Slovic 2000; cf. also, Hansson 2010 and Summerton and Berner 2003 for this category).

The third approach may be called the *cultural approach* to risk (roughly corresponding to the last two of Renn's categories). Whereas the psychological approach mainly focused on the *individual* and her ways of conceiving of risk-related affairs, the cultural approach takes a broader perspective. On this approach, the main interest is to establish how our conceptions

of risk are culturally mediated, that is, how they are formed by social contexts in our societies (e.g., identity and power). A particular risk statement is always articulated in a cultural context, singled out among many other possible formulations. Moreover, for every risk event we pick out as the interesting one, there are other potentially hazardous events we could have chosen. The cultural approach to risk is interested in, as Clarke and Short put it, “how social agents create and use boundaries to demarcate that which is dangerous” (Clarke and Short 1993, p. 79).

In this chapter, my main aim is analytical in the sense that I will analyze the conceptual aspects or constituent dimensions of risk and safety. What do terms such as “risk” and “safety” mean when we use them in statements such as “It is safe to fly,” “Nuclear power is safer than other means of energy production,” and “The risk of nuclear power is  $10^{-4}$  deaths per reactor year?” In particular, I will investigate the comparable or quantitative notion of risk and safety exemplified in the latter two statements. Here, the scientist conception of risk and safety will form an interesting starting point.

Thus, I will focus on the meaning of the relevant terms, rather than psychological and social aspects such as what individuals think about risks (a topic for risk perception studies) or how these beliefs concerning risk and what we choose as the area of study concerning risk is dependent on our social context (a topic for cultural studies of risk). That does not mean that results from psychological or sociological studies of risk are irrelevant. Since there is a close relationship between our linguistic behavior and the meaning of the term that is used, actual practice matters. In particular, I will in section [Further Aspects of Risk and Safety](#) investigate some suggestions of risk aspects that have been put forward in the literature. Furthermore, I will return to the different perspectives on risk in section [The Nature of Risk and Safety](#), where I will turn to the question about the *nature* of risk and safety, notably whether they are to be seen as objective, scientific concepts or whether they have some other status. Here, researchers belonging to different traditions have tended to take rather different stands.

## Notions of Risk

---

### Five Definitions of “Risk”

---

The terms “risk” and “safety” have been used with many related but distinct meanings. It is helpful to distinguish between at least five different, but clearly related meanings that have been used in the literature (Hansson 2004a; Möller et al. 2006):

1. Risk=an *unwanted event* which may or may not occur (Rosa 1998)
2. Risk=the *cause* of an unwanted event which may or may not occur
3. Risk=the *probability* of an unwanted event which may or may not occur (Graham and Weiner 1995)
4. Risk=the fact that a decision is made under conditions of *known probabilities* (Knight 1921; Douglas 1983)
5. Risk=the statistical *expectation value* of unwanted events which may or may not occur (Willis 2007; Campbell 2005)

The first meaning is displayed in a statement such as “Wildfires constitute the most serious environmental risk in Russia today” or “There is always the risk of an accident when driving in traffic.” “Drunk driving constitutes a major traffic risk” and “Coronary heart disease is the

number one risk of death in America” are two examples of the second meaning of “risk.” A risk in the second sense is sometimes referred to as a *hazard* in an engineering context.

The third meaning is perhaps the most common, for example, in statements of the form “the risk of heavy rainfall this week is more than fifty percent.” Note here that “risk” in this sense is not merely a synonym of “probability,” but is used to mark the undesirability of the outcome, that is, rainfall. Hence a farmer would perhaps say that it is the *chance* of rainfall that is 50% rather than the *risk* of it.

The fourth sense of risk is a technical notion that is often used in decision theory. Here, one typically distinguishes between *decisions under certainty*, which are decisions where all the consequences of the decision alternatives are known; *decisions under risk*, which are decisions where the probabilities of the outcomes are known; and *decisions under uncertainty*, where the probabilities are unknown. In this context, claiming, say, that whether to install a certain warning system is a decision under risk is to claim that the situation can be treated as a decision where the probability of failure of the system is known.

The fifth sense is another technical sense in which the notion of risk is used. While the notion of expected value as such dates from the early development of probability theory in the seventeenth and eighteenth centuries, its application in the risk context is fairly new. It became common after the influential Rasmussen Report of 1975 and is now the standard definition of “risk” in risk analysis (Rechard 1999, p. 776).

The expectation value is the probability-weighted sum of the severity of harm. It measures the *magnitude* of the risk as the combination of two factors, the probability of an unwanted event, and its severity. It supplies an overtly *quantitative* sense of risk that is used both to compare risks, and to give a single magnitude of risk.

In risk analysis, the expectation value is often used for a single quantitative statement of a risk, or as the basis of comparative statements when claims are made that some technology is safer than another. Often in statements such as “the risk of flying is less than the risk of driving a car,” it is the expectation values of both activities that are compared. (Note, however, the potential need for increased precision of such statements, for example, whether the basis for comparison is per traveled kilometer or per traveled hour.)

While all of these five meanings of “risk” are legitimate on their own terms, my main interest in this chapter is the application of the term for comparative or quantitative purposes. The last sense of the term, the notion of expected value, fits this aim. It gives a value of risk for an event that may be compared with other events, and it also gives a unique, freestanding measure of risk. In the next section, I will take a closer look at the quantitative notion, including the expected value conception.

A terminological note before I continue. In the remaining chapter, I will use “harm” instead of “unwanted event,” for several reasons. If a potential unwanted event is small, it may be incorrect to talk about risk or safety. For example, drawing a blank ticket in a lottery would be an unwanted event, but the avoidance of this would not be described as a matter of safety (unless, of course, the lottery was about something severe, such as when a person participates in a game of Russian roulette). Thus, the nature of the unwanted event is relevant here: if its severity is below a certain level it does not count as a risk and safety issue. (The term “unwanted event” also refers to the subject’s desires in an unfortunate way. If you, for some reason, had a desire to hurt yourself, and therefore engaged in a stunt act in which the probability of a severe accident is very high, we might say that you chose a *safe* way to kill yourself, meaning *certain*, but we would never say that you were *safe*.) I will therefore use the term “harm” that – contrary to “unwanted event” – implies a non-trivial level of damage.

## Acceptable Risk

---

The quantitative notion of risk that is our main object of study should be distinguished from the notion of *acceptable risk*. The quantitative notion is in place when we compare risks or ascribe their magnitude. Even if we assume an answer to the magnitude question, whether we should accept a risk of a certain magnitude is yet a further question. In other words, we should distinguish between the *magnitude* of the risk on the one hand, and whether or not a risk of that magnitude should be *accepted*. They correspond to two different notions that must be kept separated in order to avoid conceptual fallacies about risk.

Historically, early studies in risk analysis were aimed at finding a level of risk – interpreted as the expectation value of harm – that should be accepted (Otway 1987). An example of such “low levels of risk” concerns the harmful effects of background radiation, or other risks that we accept in ordinary life. While helpful as comparability tools, the idea of any such fixed levels of acceptable risk have been questioned (Peterson 2002). The general objection to the idea is that if a risk is *additive*, that is, adds to existing risks, there will be an addition to the overall risk even if the risk is small, and hence we need a further justification for adding the risk.

A major reason for accepting or not accepting a risk is naturally due to the benefit of the risk in relation to the harm of imposing it. Cost–benefit analysis of risk is a method of risk analysis that aims to judge the acceptability of a risk by comparing the benefits to the “cost” that the risk corresponds to. While the viability of cost–benefit analysis of risk is heavily questioned (Le Grand 1991; Sen 1987; Hansson 2004b; Fischhoff et al., Kirmsky and Golding, Shrader-Frechette, contributions to Asveld and Roeser 2009), among other things since it predominately relies on the controversial presumption that we may compare risk and benefits on a single (monetary) measure, it is hard to avoid the general idea that the acceptability of a risk has *some* relation to its benefits.

The acceptability of a risk depends also on many moral aspects involving questions such as agency, rights, and volition. Arguably, if a risk is voluntary, such as smoking, it may be acceptable even if the same risk, were it involuntary, would not be. Similarly, a certain risk level may be acceptable if the persons taking the risk are the ones benefiting from them, but not otherwise (Hansson 2004b).

---

## Is Safety the Antonym of Risk?

---

Risk and safety are closely related concepts. Until now, I have followed the tradition and mainly discussed the notion of risk, assuming that the relation between the two is uncontroversial. This, however, cannot always be assumed.

While risk is what we typically quantify and compare, safety is what we want to achieve. In the literature, the notion of safety is predominately used as the sought state-of-affairs. (This is the common usage in regulation and procedural documents, where the focus is on describing rules and procedures to enhance safety, but a direct and precise characterization of the concept of safety is missing; cf. e.g., IAEA 2000). We want the nuclear power plant to be safe, as we do the person walking alone in a city park and the cough medicine we use. Still, both risk and safety are used, as I have done in this chapter, to make comparative claims. The common picture of the relation between the two concepts is that they are antonyms: when the risk is low, safety is high, and conversely, when safety is low, risk is high. We could phrase a statement

“it is safer to be at sea when you are sober than when you have been drinking,” just as well as we could phrase it “it is less risky to be at sea when you are sober than when you are drunk.”

There are, however, at least three potential complications for the antonym picture of risk and safety. The first comes from the fact that the two terms have different connotations. Safety is a positive property, while risk is generally something negative. Therefore, the level of risk may perhaps not be too high for us to claim that X is safer than Y. If there is a high risk for both the plague and cholera, but slightly less for cholera, it may be strange to say that we are safer from the one than the other, even if that would be correct on the antonym view.

Secondly, one should note that the term “risk” is often – explicitly or implicitly – given a technical definition. On such a usage, it is an open question whether safety is its antonym. It has been argued, for example, that, if the expected value of harm is used as a definition of risk, safety is not to be understood as the antonym of risk, since other aspects are relevant for safety (Möller et al. 2006). On a broader notion of risk, including the aspects that will be addressed in the next section, the antonym view is more plausible.

The third complication comes from the monadic use of the safety predicate, most clearly expressible with the term “safe” – such as in “the bridge is safe” rather than the dyadic (or comparative) “the new bridge is safer than the old one used to be.” The application of the safety concept may be given an absolute and a relative interpretation. Consider the question “is my car safe?” One way of answering it is to reply: “No, since there is always a risk of being in a traffic accident – if you want to play it safe, stay home!” That way of interpreting the question would be the absolute sense of safety. According to the absolute interpretation, safety against a particular harm implies that the risk of that harm has been eliminated. Some authors take the absolute sense of safety for granted. For example, in the context of aviation safety it has been claimed that “[s]afety means no harm” (Miller 1988) and that “[s]afety is by definition the absence of accidents” (Tench 1985).

Another way of answering the question would be to reply: “Yes, the car is safe, since the risk of an accident in this car is low, and the latest safety features it comes equipped with also minimize the risk of a severe damage in case of an accident.” That reply would be in line with a relative interpretation of safety, where safety means that the risk has been reduced or controlled to a certain acceptable level. A typical example taken from a safety application states: “[R]isks are defined as the combination of the probability of occurrence of hazardous event and the severity of the consequence. Safety is achieved by reducing a risk to a tolerable level” (Misumi and Sato 1999). The US Supreme Court is explicit about its use of a relative safety concept when they claim that “safe is not the equivalent of ‘risk free’” (Miller 1988, p. 54).

The relative safety concept must not only be distinguished from the absolute interpretation, but also from the above-discussed notion of *acceptable risk*. As we have seen, whether a risk is acceptable may depend on the benefits, and on moral or social aspects. While there is a social element in expressions such as “tolerable level” of risk as well, we do not allow the same leeway to the notion of safety, and it is often reasonable to claim that while something is not safe, the risk is acceptable. Therefore, it seems as if it is actually the notion of acceptable risk that the American Department of Defense is referring to when it states that safety is “the conservation of human life and its effectiveness, and the prevention of damage to items, consistent with mission requirements” (Miller 1988, p. 54).

Both the absolute and the relative concepts of safety are legitimate, but they must be carefully separated in order to avoid misunderstandings. For most practical purposes,

the absolute concept corresponds to an ideal that cannot be realized, and “safe” is used in the relative sense. Hardly anything for which it would be interesting to apply the concepts of risk and safety contains no risk at all, strictly speaking – not even, as the first answer suggested, staying at home.

## Aspects of Risk

---

I will now look at the quantitative notions of risk and safety, that is, the concepts that are in place in statements such as “Flying is safer than going by car” or “Nuclear power has the lowest risk of all energy production methods.”

In the last definition of risk in section [Five Definitions of “Risk”](#), “risk” was defined as the expected value of harm. As noted there, this is the most common definition in risk analysis, and some authors have even taken it as the only rational way of measuring risk. Bernard Cohen, for example, starts off a recent paper with the claim (Cohen 2003, p. 909):

- ▶ The only meaningful way to evaluate the riskiness of a technology is through probabilistic risk analysis (PRA). A PRA gives an estimate of the number of expected health impacts – e.g., the number of induced deaths – of the technology, which then allows comparisons to be made with the health impacts of competing technologies so a rational judgment can be made of their relative acceptability.

In this section, however, I will investigate whether the notion of expected value of harm gives a complete understanding of risk and safety, and I will discuss other aspects that have been suggested in the literature.

Before this, however, let us start with a methodological note. It may very well be held that as long as a definition of a term, for example, “risk,” is internally coherent, it makes no sense to ask whether it is complete. We are *told* what the meaning of the term is, and if this meaning is intelligible, then no complain may be launched.

Indeed, this objection is correct as it stands. Here, however, our aim is to grasp the meaning of “risk” and “safety” in the context relevant to societal decision-making. Hence, whether the technical notion of expected value of unwanted events capture this meaning is another question. Although the expected value may supply a clear and distinct quantitative concept, the important question is whether it fits with the phenomena we are trying to capture.

In other words, our aim is to characterize the notion of risk involved in our natural language usage, not in any internal-to-science-only way. In the end, what decision-makers as well as laypeople want to know when they ask how the *risks* are distributed – or which of the alternatives is the *safest*, etc. – is the answer that is relevant for the intended meaning of the words. If we answer using the same terms but with another meaning, we have not given the adequate answer. Hence, it is important not only that we are clear about how our terms should be interpreted, but also that we are using the *right* interpretation. Risk research is principally an empirical field of study, but if we are dealing with inadequate conceptualizations, our analyses may be inadequate, too. (The approach used here has much in common with *ordinary language philosophy*, in which a core assumption is that knowledge of the meaning of terms is reached not primarily through theories in abstraction, but through a close attention to the details of ordinary language in which it is used (Soames 2005). Cf. also Hare (1952, p. 92) for a similar point against the claim that we may define moral terms as we please.)

## Harm and Probability

---

The notions of harm and probability are central for any quantitative understanding of risk and safety. As soon as we move beyond an absolute conception as discussed in section [Is Safety the Antonym of Risk?](#), where safe means no risk at all, we are dealing with events that *may* take place with different degrees of likelihood, and causing different levels of harm.

Starting with harms, we should note that it is far from trivial to compare their severity. On a commonsense notion of harm, it seems clear that there are many cases in which we may reasonably hold that one harm is more severe than another. In a traffic accident, for example, a death is more severe than a broken leg, and a broken leg is more severe than a bruise. But the relative severities between them are not so certain. How many broken legs, one may reasonably ask, are there on a death? It is not clear what the answer should be.

This is problematic if we think that we can always, in principle, compare risks. In particular, for the expected value of harm to be well defined, the severity of harm must be possible to measure rather extensively. More precisely, we must be able to decide not only that harm A is more severe than B, and that B is more severe than C, but also the relative severity between them. (In technical terms, we must be able to compare harms on an interval scale; cf. Resnik (1987) for a textbook presentation.)

In practice, when arguing for (and applying) the expected value notion of risk one typically limits the measure of the severity of harm to one significant value, typically casualties. So severity of harm in traffic is in the first instance measured in traffic deaths. A perhaps sometimes reasonable assumption behind such a measure is that the less severe harms “follow along,” but that is sometimes a far-fetched idealization.

Even if we assume that death is the primary harm to take into account, it is unclear if the severity of harm is the same in all cases. Is, in all circumstances, the possibility of a 95-year-old person dying from a medical procedure of the same severity as if the patient were 25 years old? This is an area of severe controversy, but we should note that there are ambitious systems of measurement designed to take account of both the quantity and the quality of life generated by healthcare interventions, such as the QALY measure (quality-adjusted life-year).

The second aspect of risk, probability, is a much-investigated concept with a well-established mathematical content, and it is paramount in many important applications. The interpretation in cases of risk and safety is not evident, however. In probability theory there is a well-established distinction between subjective and objective interpretations of the concept. According to the objective interpretation, probability is a property of the external world, for example, the propensity of a coin to land heads up. According to the subjective interpretation, to say that the probability of a certain event is high means that the speaker’s degree of belief that the event in question will occur is strong (Ramsey 1931; Savage 1972). When we are dealing with the repetition of technological procedures with historically known failure frequencies, it may be possible to determine probabilities that can be called objective. However, in most cases such frequency data are not available, unless perhaps for certain parts of the system under investigation. Therefore, frequency data will have to be supplemented or perhaps even replaced by expert judgment. Expert judgments of this nature are not, and should not be confused with, objective fact. Neither are they *subjective* probabilities in the classical sense, since by this is meant a measure of a person’s degree of belief that satisfies the probability axioms but does not have to correlate with objective frequencies or propensities (see section [Uncertainty](#)). The judgments are better described as subjective estimates of objective probabilities. Furthermore, the probability

estimates used in risk and safety analysis are not purely personal judgments even in this sense. Rather, they are based on the best possible judgments that can be obtained. Typically, a probability estimate is interpreted as the best possible judgment from the community of experts.

The theoretical issues of comparing harms and assigning probabilities do not only constitute a problem for proponents of the expected value notion of risk, but for any account that, while objecting to the expected value interpretation *as such*, still takes severity of harm and probability as basic aspects of risk and safety. To the extent that there is a problem in assigning a reasonable probability or severity of harm value for an event, there is a potential problem in assigning a measure of risk. As will be seen in the next two subsections, this problem is present even if we manage to agree about an interpretation of the most reasonable ascription of the probability and severity of harm that can be obtained.

## Risk as the Expected Value of Harm

---

That the severity of harm and its probability are important aspects of risk is well in line with basic intuitions about the notions. Assuming that the measurement problems in the last subsection are solved, it seems reasonable to assume, *all else equal*, that if the likelihood of an accident increases, so does the risk; likewise, if the severity of an outcome increases, so does the magnitude of the risk. So far, it is hard to deny that these two aspects should be part of the quantitative concepts of risk and safety. That, however, falls short of accepting that the expected value of harm supplies the correct measure of risk.

Indeed, some authors, while endorsing the two aspects, have preferred more cautious formulations. Slovic, for example, has defined risk as “a blend of the probability and the severity of the consequences” (Slovic 2000, p. 365). Other authors have likewise expressed themselves in terms of “combinations” rather than explicitly stating how that combination is brought into a measure (Lowrance 1976; ISO 2002; Aven 2007). Still, as previously noted, the expected value is the dominant notion of risk in risk analysis (Hansson 2005).

There is, however, a strong reason for the claim that the expected value of harm is the proper measure of risk. It is, proponents argue, the only measure the use of which will minimize total harm *in the long run*. If we assume the *expected* number of people dying of heart attack in a country is 400,000 per year, that typically means that about 400,000 *will* actually die from heart attack on a given year. Imagine that we had a new revolutionary treatment for coronary heart disease that would decrease the expected number by 100,000. Switching to the first treatment would typically mean that around 300,000 people would die the following year.

This strong correlation between the *expected* value of an outcome and the *actual* value follows from what is called the Law of Large Numbers, which is the theorem in probability theory that states that if independent trials with the same probability of outcomes are repeated, the average value of the trials converges to the expected value. In other words, in the long run, given the above assumptions, actual value converges to the expected value. The classic illustration of this is the tossing of a (non-biased) coin: whereas there is nothing strange in tossing three heads in a row, we would expect the relative frequency after, say, 100 or 1,000 tosses to be very near to 0.5.

There are, however, several objections to defining the risk as the expectation value of harm. In the previous section I discussed some problems with comparing harms in a sufficiently extensive way as to make the notion of expected value well defined. Even if we assume that such a measure is available, however, there are problems with defining risk as expected value.

One main objection takes on the background assumption that is explicit in the very name of “Law of Large numbers.” It is based on noting that as soon as the numbers of events in question are less than “large,” there is no guarantee that the actual outcome is close to the expected one. (Indeed, *given* any number of events, there is no guarantee that the expected value is close to the actual outcome.) Let us imagine a future where we have a much smaller number of people dying from heart attack, say typically 20. In this future society, we are faced with a clear choice (don’t ask me about the mechanisms for this, it is a thought experiment) between medicine A or B. With medicine A there is a 0.00001 probability that 1,000,000 will die, and 0.99999 that no one will die. With medicine B there is a probability of 0.5 that no one will die, and 0.5 that 30 will die. With medicine A, the expected value is 10, whereas the expected value with medicine B is 15. The proponent of the expected value notion of risk would thus claim that using medicine A would mean the least risk, but this, the objectors argue, is not at all clear. The *actual* outcome could in fact be 100,000 deaths in heart attacks the following year, if treatment A is chosen, whereas in case B, it would be, at maximum, 30 persons. Hence, they may conclude, the safest option is B.

The basic idea in this objection is thus that there may be a discrepancy between the expected value and the actual value, depending on the actual distribution of outcomes and their matching probabilities, and that it is thus not irrational to oppose the general identification of risk with the expected value of harm. If we are dealing with “one-shot” events rather than frequent events of the same type and the same independent probabilities, the expected value may be nothing more than *one* aspect of the risk, not the full determinant of its magnitude. In particular, when the potential harmful outcome is extreme, such as the extinction of mankind, the expected value notion is deficient as a measure of the risk.

## Uncertainty

---

While probability and harm are often alluded to as the basic aspects of risk and safety, some authors explicate the terms by focusing on the aspect of *uncertainty*. Hence, risk is sometimes explained with reference to an uncertain consequence of an event (Renn 2005; Aven and Renn 2009). Under some interpretations, however, this notion of uncertainty reduces into probability. Aven (2003: xii), for example, uses probability and probability calculus as “the sole means for expressing uncertainty.” This probabilistic understanding of risk and safety has a substantial justification in the subjectivist *Bayesian framework* (Howson and Urbach 2006). As mentioned above, on a subjectivist interpretation, probability is conceived of as representing all aspects of a decision-maker’s lack of knowledge. On a Bayesian construal, *all* rational decisions are fully representable with precise probabilities, since the rational decision-maker always, at least implicitly, assigns a probability value to each potential outcome. Faced with new information, the agent may change her probability assessment (in accordance with Bayes’ theorem), but she always assigns determinable probabilities to all states of affairs. Thus, in the Bayesian view all uncertainty about what will happen is codified in the probability assessment for the outcome at hand (Ramsey 1931; de Finetti 1937; von Neumann and Morgenstern 1944; Savage 1972).

The reduction of uncertainty to probability is however not the only alternative. On the contrary, as was argued above the probabilities behind risk and safety ascriptions are not typically given a subjective Bayesian interpretation. Rather, it is more in line with what is called

the classical view in decision theory, in which a basic distinction is made between situations with known probabilities and situations where probabilities are unknown or only partially known (Knight 1921). Proponents of the classical view have pointed out that there is a large difference between situations with well-determined probabilities, such as coin-tossing, and less well-determined situations such as assigning probabilities to whether a major accident will happen in a complex plant. In the latter case there is *epistemic uncertainty* that, according to these authors, may not be reducible to a unique probability value in a rational way. This has led many contemporary theorists also of Bayesian bent to argue for the inclusion of (non-probabilistic) epistemic uncertainty into the analysis (Ellsberg 1961; Kyburg 1968; Levi 1974; Gärdenfors and Sahlin 1982).

One way of illustrating the idea of epistemic uncertainty is to imagine two situations in which you are walking alone in the jungle, and are about to cross an old wooden bridge. The bridge looks unsafe, but you have been told by people in the local village that the probability of a breakdown of this type of bridge is less one in ten thousand. Contrast this with a case where you in your jungle walk are accompanied by a team of scientists, come across a bridge, and the scientists carefully investigate the bridge and conclude that the probability that it will break is rather one in five thousand. Even though the probability is now judged as higher, the epistemic uncertainty is far smaller and it is not unreasonable to regard this situation as preferable to the first one in terms of safety.

This kind of example suggests a picture in which the risk increases with the probability of harm, with the severity of harm, and with uncertainty. So two events with the same severity of outcome that are given the same probability for occurring in accordance with the best estimations possible may still warrant different ascriptions of risk, if the uncertainty is considered different in the two cases. Assume, for example, that we are to compare the safety of two different means of energy production that are both given the same probability of severe failure, one utilizing an old and well-tested method, and another utilizing a new method. Naturally, the new method has been well tested as far as possible, but we do not have the same source of performance data for the entire system as in the case of the old method. Here, it may be reasonable to claim that the risk of the new system is higher, due to the greater epistemic uncertainty in question.

Many methods and techniques in safety engineering may, in the first instance, reasonably be interpreted as methods of reducing the uncertainty. Take the method of *inherently safe design*. Here the idea is to switch from controlling potentially dangerous materials and subsystems to using materials and subsystems that are less dangerous in themselves, that is, that do not need to be controlled in the same degree. Hence, a possible hazard is *removed* rather than *contained*. For example, fireproof materials are used instead of inflammable ones, and this is considered superior to using inflammable materials but keeping temperatures low (Möller and Hansson 2008).

While the general notion may be reasonably clear, a fundamental question is how epistemic uncertainty should be characterized in more detail. This is a controversial area in which no consensus has been reached. The most extensive discussions have been in decision theory regarding how to express the uncertainty of probability assessments. Here, two major types of measures have been suggested, binary and multi-valued measures. A binary measure divides the probability values into two groups, possible and impossible values. In typical cases, the set of possible probability values will form an interval, such as: “The probability of an epidemic outbreak in this area within the next 10 years is between 5% and 20%.” Binary measures have been used, for example, by Ellsberg, Kaplan and Levi (Ellsberg 1961; Kaplan 1983; Levi 1986).

Multivalued measures generally take the form of a function that assigns a numerical value to each probability value between 0 and 1. This value represents the degree of reliability or plausibility of each particular probability value. Several interpretations of the measure have been used in the literature, for example, second-order probability (Baron 1987; Skyrms 1980), fuzzy set membership (Unwin 1986; Dubois and Prade 1988), and epistemic reliability (Gärdenfors and Sahlin 1982). See Möller et al. (2006) for an overview.

In summary, while many authors agree that epistemic uncertainty is an important aspect for risk and safety *in addition* to probability, how to include it in more detail is strongly controversial. Perhaps this should not surprise us, however. A general measure of what we do not know may sometimes be too much to ask for, even in theory.

## Further Aspects of Risk and Safety

---

For the quantitative notion of risk, several candidate aspects beyond probability, severity of harm and epistemic uncertainty have been suggested in the literature. In particular, empirical studies of risk – so-called *risk perception studies* – have established many other aspects that are relevant for how people judge risks. For example, in a seminal study Fischhoff et al. (1978/2000) presented eight aspects that determined people's preparedness to accept a risk. In addition to severity of consequence – which I have already discussed – the aspects were: Voluntariness of risk, Immediacy of effect, Knowledge about risk, Control over risk, Newness, Chronic-catastrophic, Common-dread. Results like these have helped shape a demand, directed at risk assessors, to take other aspects than the traditional under consideration.

Now, as I noted in section  [Acceptable Risk](#), there is an important distinction between the *magnitude* of a risk and its *acceptability*. Many of the aspects in the above study seem relevant to the *acceptability* of risk. Still, these aspects return as suggestions for aspects relevant to the very notion of risk as well. May there be any merit to such suggestions?

In evaluating such a claim, we should first distinguish individuals' *conceptualization* from the *concept* in question. While there typically is a deep connection between competence and actual use of a concept, many of our beliefs about a concept can be wrong even if we are competent with it. Moreover, we seldom think that we have full knowledge about how to exactly delimit a certain concept. I take myself as pretty competent with a rather wide range of concepts – such as *table*, *computer*, *book*, and *coffee mug*, just to mention some of the ones I just applied to objects in my proximity – but I am not sure that I would be able to delimit any of these concepts with any certainty. So even if we did show that there is an aspect that many people *take* to be part of the concept of risk, it does not mean that it actually *is*.

Such a skeptical stance has been a common attitude for many traditional risk analysts faced with suggestions beyond their preferred notion of risk, that is, typically the expected value notion (Hansson 2010). In their arsenal, they have used risk studies that seem to have shown the irrationality of people's risk perception. For example, it has been shown that people often take the risk of exposure to chemicals and radiation as something binary – either you are exposed or not – and neglect the dose factor (Kraus et al. 2000/2000). Since actual harm is predominately a factor of dose, it amounts to a misconception of the risk involved.

It certainly is a radical claim that such beliefs about chemical exposure are *irrational*. It is indeed an intelligible position – it just happens to be wrong. But false beliefs are not irrational beliefs. The point to make here, however, is that while there is a *prima facie*

reason to take strongly held beliefs about a notion seriously, if we have sufficient evidence we may show that they are mistaken.

When it comes to many of the suggested candidate aspects of risk and safety, I believe indeed that they are mistaken, when thought of as *conceptual* aspects. Take the aspect of *control* (Möller et al 2006). To say that a certain risk is *more controllable* for an agent than another risk means, in the present context, that there is a more reliable causal relationship between the acts of the agent and the probability and/or the severity of the harm (this is an extension of the psychometrical concept, which only regards the subjective dimension). At first sight one might believe that, everything else being equal, more controllability implies less risk. Arguably, if we only have the option of flying on automatic pilot, we would be less safe than we would if we had the option of turning it off as well. However, this relationship does not seem to hold in general. We have, again, to distinguish between a psychological perception of safety and actual safety, since we may feel safe without actually being safe, and *vice versa*. Consider two nuclear power plants, one of which runs almost automatically without the intervention of humans, whereas the other is dependent on frequent decisions taken by humans. If the staff responsible for the non-automatic plant is poorly trained, it seems reasonable to maintain that the automatic plant is *safer* than the non-automatic one *because* the degree of controllability is lower in the automatic plant. Arguably, even if they are excellently trained, a high degree of control may lower the safety. Consider a third nuclear plant that is much less automated than the ones we have today. Due to cognitive and mechanical shortcomings, a human being could never make the fast adjustments that an automatic system does, but this plant has excellent staff that perform as well as any human being can be expected to do under the circumstances. In spite of this increased control this too would be a less safe plant. The reason is that the probability of accidents due to human mistakes would be much greater than in a properly designed, more automatic system.

The effects of control on safety in our nuclear plant example can be accounted for in terms of the effects of control on probability. If increased human control decreases the probability of an accident, then it leads to higher safety. If, on the other hand, it increases the probability of accidents, then it decreases safety.

Denying that control is part of the *concepts* of risk and safety is not denying that it may have other relations to the phenomena. For example, it may be a *heuristic device* in many situations: it might be that having the option to causally control our systems often would increase their safety. This correlation, however, is an *empirical* possibility rather than a conceptual one.

For other aspects, it seems rather clear that what is at stake is not the quantitative notion of risk and safety at all, but that they are reasons for the acceptability of a risk. Voluntariness of risk, for example, seems to be about the acceptability of imposing involuntary risk and not a claim that the degree of voluntariness changes the magnitude of risk in any sense.

There are, however, aspects of a potential risk event beyond probability, severity of harm and epistemic uncertainty that have been argued to belong to the concept of risk and safety (Möller 2009a, 2010). Although admittedly controversial, they are instructive in order to further demonstrate the complexity of inferring ascriptions of risk and safety.

*Distributive aspects* illustrate a further complication with harms and risk: even if we assume a measure, such as a reasonable way of comparing a broken leg with a casualty (as discussed above), how to infer the risk from different *distributions* of potential harms may not always be clear (Hansson 2005; Möller et al. 2006).

Imagine that a revolutionary method of building encapsulated nuclear power plants has made the population safe from even the unlikely event of a meltdown and thus decreased

the total expected value of harm. The only drawback is that service staff must make some internal maintenance and this is very risky, having a high expected value of harm, several magnitudes above current levels for any staff. Let us further assume that the level of epistemic uncertainty is deemed to be negligibly low in both the current and the new and revolutionary method.

The question we can ask ourselves is whether, in cases like these, safety is merely a matter of receiving as low an expected value of harm as possible in a population, or if the distribution of potential harms also should count? Maybe we should not hold that a situation is safer than another if there are some people carrying significantly higher risks than others? In most societal activities carrying risks, there is an uneven distribution of potential harm. Persons living close to risky artifacts such as energy production plants or heavy chemical plants take a higher risk load than persons living further away. Likewise, for road traffic safety there is a debate about using cables for protecting the vehicle from driving off the road. The cables are successful in avoiding many potentially lethal car accidents; however, they may also be very dangerous for motorcyclists; much more so, in general, than if the motorcyclists merely went off the road. Since there are many more cars than motorcycles on the roads, the expectation value of harm is smaller when using the wires than not, yet for the motorcyclists the risk is much higher. It may be asked whether it is then reasonable to claim that the method is safe as long as the total expected value is kept low, regardless of the particular risks for motorcyclists.

*Delimitation issues.* The second aspect I will mention highlights the complicated nature of selecting the base events for ascriptions of risk and safety – in this case the harmful potential outcomes (Möller 2010). Let us imagine a situation where we are to judge whether automobile traffic between two particular cities is safer than airplane traffic. Extensive frequency data tells us that the probability of serious harm when traveling by airplane is lower than traveling by car. We have reason to believe, however, that to a significant extent, the one-car accidents are intentional, that is, acts of suicide. For air traffic, however, there are almost no accidents that may reasonably be labeled as suicide. If we were to exclude the suicide cases from the frequency data, the expected value of harm for car travel would be lower.

In this thought example, we face the question of how to delimit the risk and safety concepts: what events should count as risk events, and hence be included in frequency data and other means of evaluating the risk and safety at hand? If we were to include all harmful events, air travel would be safest. But it seems wrong to include also suicides in the statistics of road traffic safety. It is one thing to allow mistakes from the driver – and a main part of accidents certainly derive from the human factor – but in judging what means of transportation is the safest, it seems irrelevant that it is possible to use the car also as a tool of committing suicide. Note that this is not a question of epistemic uncertainty (although in practice there certainly is such a question as well). The question is rather that given that we know that an event is a traffic suicide event, should we include it in the basis for the frequency data? It is evident that only some harms count, not others, and we may reasonably doubt the answer, in the thought experiment, to which means of transportation is the safest.

Admittedly, in these two cases of distribution and delimitation, our intuitions may vary. But it seems at least plausible to doubt here that what has the lowest expectation value of harm in these cases are actually safest. Note that for this to be plausible, we do not need to claim that the actual risk of one event is greater or even equal to the other when the expected value of harm of it is less. The weaker claim that the risks are incommensurable, that we cannot reasonably compare them in such a case, is sufficient.

As mentioned above, there are many aspects of risk and safety proposed in the literature; distribution and delimitation aspects such as mentioned here are mere examples of the questions involved in inferring the risk from a given, complex situation. Still, they exemplify the possibility of further reasonable inferential aspects for the quantitative notion of risk, even beyond the general question of how the dimensions of probability and severity of harm should be combined into one measure. In the next section, we will investigate whether these aspects, among others, represent a problem for a scientific notion of risk.

## The Nature of Risk and Safety

---

The nature of risk and safety has been understood in widely different ways in the literature. Some authors have taken risk to be a scientific, objective, or natural notion, picking out real properties in the world, while others have taken it to be something far more normative or subjective, or even denied its existence. Generally, these commitments can be mapped to a high degree with the three risk perspectives that were mentioned in section [Risk Perspectives](#): the *scientist notion* of risk, the *psychological approach*, and the *cultural approach*. On the scientist approach, risk is typically seen as a scientific notion and, correspondingly, a phenomenon that may be investigated and measured in a systematic way, at least in principle. With that often comes the idea that risks are natural, objective, and real features of the world. On the *psychological approach*, the focus is on how individual people conceive of risks and their acceptability. But many psychological researchers are interested in how this relates to how the risks actually *are*, and often seem to assume that there is an objective fact of the matter to compare with (Summerton and Berner [2003](#), p. 6; Hansson [2010](#), p. 232).

The strongest opposition to conceiving risk and safety as objective concepts comes from proponents of the *cultural approach* to risk. As mentioned in section [Risk Perspectives](#), on the cultural approach, risk is typically seen as a subjective and social phenomenon. Often this is expressed in terms of a social construction: risk is a social construction, since the content and delimitation of risks are socially articulated and treated. People from different cultures as well as within one culture have very different views on what constitutes a risk and how severe a risk may be. Sometimes proponents of this view express that there is no fact of the matter over and above these individual or cultural views, committing them to a denial of risk as objective and perhaps even real. (These are merely initial characterizations: as will become clear, all of the three risk perspectives are *compatible* with objectivist, subjectivist, or constructivist conceptions of risk as described below.)

Even if we assume that there is a lot to say in favor of the initial characterization just given, it should be unpacked a little more carefully. Taking risk to be *natural*, *objective*, or *real* are distinct thoughts that I will investigate further in the following section. First, in section [Risk and Normativity](#), I will distinguish between taking risk and safety as *natural* concepts on the one hand, and as *normative* concepts on the other, and discuss some allegedly normative aspects of risk and safety. In section [Thick Concepts and Reductionism](#), I will present an interpretation of risk and safety as what in philosophy has been called *thick concepts*, and discuss the idea that risk can be reduced to a natural concept. Lastly, in section [Risk, Objectivity, and Social Constructions](#), I will discuss the idea that risk and safety are social constructions, and the related question whether that means that risks are not objective or real.

## Risk and Normativity

In the traditional use, a natural concept is a concept that is invoked in scientific explanations or, more narrowly, used to express the laws of nature (Little 1994; Vallentyne 1998; the latter characterization is in line with traditional understanding of natural kinds: the close relation between natural kinds and law-like regularities is emphasized by most natural kind theorists, cf. Lange 2007; Boyd 1991; Hacking 1991; Wolf 2002). The philosopher G.E Moore has offered a number of influential characterizations to this effect, suggesting that a natural property is a property that is the subject matter of natural science and psychology, and that it is a property that can be known by means of empirical observation and induction (Moore 1903, pp. 25–27). Paradigmatic natural concepts are *water*, *gold*, *mass*, and *redness*.

Normative concepts are typically characterized as opposed to natural concepts. An often-used image is that the natural and the normative have opposite “directions of fit”: to be true, natural claims should fit what the world is like, whereas it is what the world is like that should fit the true normative claim (Williams 1985, Blackburn 1998). Normative concepts such as *good*, *right*, and *fair* are thus action-guiding; that an action is good, right, or fair entails that we have a reason to perform it. (I am in this chapter using “normative” broadly in a sense that refers both to deontic concepts such as *right*, *ought*, and *permitted*, and to value concepts such as *good*, *bad*, and *better*. Sometimes “normative” is used to refer only to the former category, while “evaluative” is used only for the latter. While the relation between deontic and value concepts is an interesting topic in its own right, for the purpose of the present chapter, it is the distinction between the cluster of these broadly normative/evaluative concepts on the one hand, and the natural concepts on the other, that is of interest.) In contrast, natural concepts have no such *prima facie* action-guiding feature: that something contains water may be a reason to drink it if you are thirsty, but not otherwise.

Is risk a natural or a normative concept? As we have seen, many proponents of the scientist approach to risk view it as a perfectly natural, descriptive concept. This goes in particular for proponents of the expectation value of harm conception of risk, where risk is understood as a function of probability and severity of harm, where both harm and probability are taken as natural concepts.

However, the picture of risk and safety as natural, scientific concepts has been criticized. The criticism is grounded on either *external* or *internal normativity*.

*External normativity* is the normativity involved in answering questions about whether or not a risk should be imposed, is acceptable, or small enough to be safe. I label it “external” since it is external to the quantitative notions of risk and safety that have been the main subjects of investigation in the current chapter.

Many sources of external normativity have already been mentioned in section [● Acceptable Risk](#) on acceptable risk. As soon as we leave the realm of absolute safety as discussed in section [● Is Safety the Antonym of Risk?](#), where safety is interpreted as a level of no risk, the question whether a certain level of risk should be accepted, or deemed safe, is relevant. In the early studies of acceptable risk, much work centered on what level of risk people thought was acceptable. Reasonably, what people *take* to be an acceptable level of risk is relevant for what level we *should* accept. Indeed, this is a basis for democratic decision-making. But generally they are different questions. The latter question about what level we should accept is a paradigmatic normative question. Hence, even if we take a certain level of safety as a given,

whether that level is *sufficiently* small to be acceptable or safe is a further normative question about what we have reason to do.

This normative status is most obvious when the reasons for accepting or not accepting a risk do not have to do only with the levels of risk as such – e.g., whether they are larger than other risk levels that we have accepted (see section 2 [Acceptable Risk](#)) – but with moral considerations, such as rights, agency, and autonomy. It is particularly evident when the acceptability of a risk depends on moral reasons that go in different directions. For example, that the direct risk of smoking is voluntary may be seen as a reason for allowing it in public areas. On the other hand, there may be a statistical connection between allowing people to smoke in public areas and the number of people that actually take up smoking. That could be considered as a reason against allowing it. The outcome of such a question of whether or not to allow public smoking, however based on scientific results on its harmful effects, is clearly a normative statement.

*Internal normativity.* That questions of acceptable risk and safety are normative, however, is hardly controversial today, especially after the evidently normative problems involved in trying to set a general limit for acceptable risk mentioned above. To the contrary, in modern risk analysis one is careful to distinguish between the quantitative assessment of risks with the assessment of the viability or acceptability of the risk. This corresponds to the typical division of risk analysis into two stages: *risk assessment* and *risk management* (NRC 1983; EC 2003; sometimes, as in the NRC model, risk assessment is divided into two stages as well, the research stage and the assessment stage). Risk assessment is traditionally considered the scientific stage where the estimations of the risks at hand are produced. Risk management then uses the output of the risk assessment as input for making a decision about the risk, ultimately whether to accept or reject it.

Thus, theorists that claim that risk is a natural concept – who we will refer to as *naturalists* – should typically be seen as referring to the *quantitative notion* of risk. Compare with paradigmatic measures such as *length*. Here, we may claim that someone's length may be a natural, scientific property and still agree that whether that length is acceptable for a certain task, such as being a basketball player, is a normative affair.

Proponents of a naturalist view on risk have several reasons in their arsenal. While the nature of probability is arguably a debated subject, as previously mentioned what is normally of interest in risk and safety contexts is not a subjective Bayesian notion but rather an objective notion such as relative frequencies (Resnik 1987). In addition, it is often claimed that severity of consequences may in many cases be measured to a sufficient degree. A basic measure is the number of lives at stake, so that higher severity means more potential deaths.

Critics, on the other hand, point to the very complications concerning the aspects of risk that I have discussed in this chapter. Probability, for example, may seem as an untouchably scientific notion due to its profound mathematical basis, but while a relative frequency is indeed a statistical fact, its role in deciding the actual probability of an event may be questioned, since it records historical data and the question is always how that applies to the situation at hand. Moreover, for many complex systems, we have sufficient frequency data for only parts, and have to estimate the overall probability, and there may be no value-free way of making that estimate.

This dependence on values is even clearer when we move beyond probability to epistemic uncertainty. There is no established measure of uncertainty. The many alternative suggestions put different aspects of uncertainty to the forefront, and it is unclear that there can be a general measure of such an elusive entity.

Naturalists can reply that these objections merely point to the fact that there are *epistemic values* involved when making risk and safety ascriptions. However, they may continue, this is not anything specific to the risk and safety area – in fact, philosophers of science have argued that epistemic values such as reliability, testability, generality, simplicity, etc., are integral to the entire process of assessment in science (McMullin 1982; Kuhn 1962; Lakatos and Musgrave 1970). Hence, they are the kind of values that the scientific experts are competent to apply in their scientific enterprise. Naturalists conclude that showing that risk ascriptions are normative in this sense merely puts it on the same level as our most paradigmatic natural notions such as *water* and *mass*.

In other words, on the naturalist defense, we have to be skeptics about *all* scientific notions in order to be skeptical about the notion of risk. Some theorists bite the bullet and draw that skeptical conclusion (Mayo 1991, pp. 254–255; Wynne 1982, p. 139). But even if we accept that all scientific claims are conditioned by epistemic values, and that this does not make them normative notions in the sought sense (merely an internal-to-science-only sense), it may be doubted that the values involved in risk ascriptions are limited to these values. The values involved in criteria for accepting a theory or an empirical result into the scientific corpus are typically biased toward avoiding false-positives, that is, taking something as a scientific fact when it is not. For risk and safety analysis, however, we may have a broader scope in order to avoid underestimating the risks (Möller 2009b; Wandall 2004).

Probability and epistemic uncertainty are only two dimensions of risk and safety, however, and the problem for the risk naturalist may be even larger when we turn to harm, and to the further problem of inferring risk and safety from all of these aspects. Even if we grant that a certain harm may be a perfectly natural notion, it is less clear that severity of harm may be ascribed in any value-free way. As I mentioned in section [Further Aspects of Risk and Safety](#), neither comparing different harms nor ascribing a level of severity to an event such as a death seems to be possible to do in a value-free way. Indeed, it seems normative if anything is.

Furthermore, there is the added complication of inferring risk and safety from its “base aspects.” As we have seen, the traditional risk assessment measure of expected value of harm has been heavily criticized. If we do not accept the expected value as a measure of risk, how then are we to infer the risk? Is there a scientific way of making this inference, one using – at the most – only values internal to science? The skeptic to risk as a natural concept doubts this.

If we include also the more controversial inferential *distribution* and *delimitation* aspects discussed in section [Further Aspects of Risk and Safety](#), the risk naturalist faces even further obstacles. To argue – as in the case of the nuclear power plant where staff were to be exposed to a significantly larger potential harm than the population at large – that there are cases where the uneven distribution of potential harm is relevant for the very level of risk (and not only for the acceptability of a certain level) is to argue that distributive aspects are *part* of the very concept of risk and safety. And it seems doubtful, even if we grant that natural science may give us a probability and a severity of harm for an event, that there is anything for natural science to say about such distributive questions as these.

The delimitation aspect also seems to be a normative aspect of risk and safety, that is, a normativity due to which events should count as safety events, relevant for assessing the question of the risk and safety in the case at hand. The traffic suicide case indicates that not *all* potential harms are relevant for risk and safety. What is qualitatively identical on the level of physical harm is not necessarily identical for evaluating the risk and safety because one event is a suicide and the other an unintentional accident. It has been argued (Möller 2009a, 2010) that

there are several potential properties of a harmful event that may motivate an exclusion of it as a risk-relevant event in certain circumstances, e.g., that it constitutes harm-seeking behavior or harms that are too small in the relevant circumstance (even if they are frequent), etc.

## Thick Concepts and Reductionism

---

As we could see in the subsection above there are several arguments for the normative nature not only of notions such as *acceptable risk* and *safe*, but also for the normativity of the quantitative notion of risk itself. Is risk as a normative notion a threatening idea for the proponent of the naturalist notion of risk? Perhaps the naturalist may acknowledge that risk and safety are normative notions and yet argue that they may be reduced to natural ones? After all, it should be clear that the notion of risk has a substantial amount of non-normative content. In this section, I will place this idea in a broader philosophical context, namely, the philosophical debate of thick evaluative concepts.

When I introduced the notion of normative concepts above, I mentioned such normative concepts as *good* and *right*. Even if the arguments for risk and safety as normative concepts are sound, they seem to differ from these paradigmatic ones. That an action is good or right tells you that it is the thing to do, at least *prima facie*, but it does not tell you anything about what type of action it is. It does not have any (or at least not much) descriptive content, only normative content. To be kind to a stranger may be morally right, but so can, many believe, starting a war (that is the assumption behind the “*just war*” concept). And these actions seem to have very little in common, except for the – assumed – normative status of being right.

Certainly, there is also a normative dimension of risk. That something constitutes a risk is typically a reason against (allowing, using, performing) it, and the larger the risk, the stronger is that reason. Risk has a negative evaluative “direction,” just as safety has a positive one. But risk is also importantly different from the paradigmatic normative notions of right and good, in that it is a notion with substantial descriptive content. Indeed, the richness of descriptive content is the main argument for the scientist approach to risk and safety. There are many descriptive characterizations of risk and safety that render them quite different from the paradigmatic normative notions of good and right: the central role of severity of harm; that a harm is not certain but may obtain; that a risk is greater the more likely it is that the harm may occur and the greater the harm, etc. Perhaps, proponents of the scientist approach to risk may argue, there is a possibility to acknowledge that risk has a normative aspect to it, but that we may, for scientific purposes, *reduce* the notion to its natural part.

In moral philosophy, there is an analogous debate among the kind of concepts that are called *thick normative concepts* – or thick concepts for short. In philosophy, John McDowell and Bernard Williams introduced the notion of thick concepts in the 1970s and 1980s (McDowell 1978, 1979, 1981; Williams 1985). Thick concepts are concepts such as *cruel*, *brave*, and *selfish*, concepts that have *both* descriptive and normative content. As such, thick concepts seem to differ from both paradigmatic natural concepts such as *water* and *length*, which have no such normative quality, as well as from paradigmatic normative concepts such as *good* and *right*, which are said to be solely or primarily normative. On the traditional analysis, a thick concept fills a double function: it describes a feature (world-guided function), and it evaluates it (world-guiding function).

On our characterization of risk and safety, it seems as if these notions fit right into this category: that something is safe is a positive feature of the entity, and that something carries a risk is a negative feature of it. But it is not simply positive or negative, it is positive or negative *in a certain way*; it has a certain descriptive “shape,” as I noted above. Grasping these aspects of the concepts is part of what one has to do in order to understand their meaning. Hence, it has been argued, risk and safety are to be understood as thick evaluative concepts (Möller 2009a, 2010).

When theorists first started to investigate (what would later be called) thick concepts, they believed that these could be analyzed as a *conjunction* of a descriptive part and an evaluative part, by which is meant that the descriptive content and the evaluative content can be independently given (Stevenson 1944; Hare 1952). “X is courageous” could therefore be analyzed as something along the lines of “X intended to act in the face of danger to promote a valued end” and “this is (*prima facie*) good-making.” Furthermore, it was commonly assumed that the descriptive part determined the reference of the thick term. A thick concept, on this analysis, is a concept with a descriptive shape that decides the extension, and an evaluation that commends or condemns an entity for having these descriptive features. If this were correct, a descriptive reduction of a thick concept would, at least in principle, be possible. While something may still be said to be lacking in the understanding of the concept of courage in a person who grasped the descriptive part but took a neutral evaluative stance toward it (did not believe that courage was either good or bad), this person would be fully able to identify courage. For the concepts of risk and safety, this would mean that there indeed was a descriptive part that one could – in principle – isolate and operationalize. Perhaps an agent that understood only the natural part of risk and safety would not have a full understanding of the concept and all of what it signifies, but she would still be able to get the extension of the concept right. If this is possible, risk and safety would be normative concepts, but they would still be able to be *reduced* to natural concepts in the following important sense: natural concepts would still suffice as the descriptive kernel of the concept that would allow us to compare risk in a scientific, non-normative way. The normativity of risk and safety, on this understanding, could then be acknowledged without any threat for the scientific status of risk ascriptions.

To make this reductive idea clearer, it may be helpful to compare with the notion of body mass index (BMI), which is an individual’s body weight divided by the square of his or her height). Having a BMI higher than 25 (based on the standard unit values kilogram and meter) is typically evaluated negatively, as it is considered bad for the health. Conversely at the other end of the spectrum, a BMI below 18.5 is considered a too low relative weight. In this sense, BMI may be seen as a concept that is a compound of both descriptive and evaluative parts. Here, however, the descriptive part is sufficient for picking out the extension of the term.

BMI is an example of a term with fully technical primacy: it is developed and applied using only natural notions. Hence, it is not at all surprising that these natural aspects are sufficient – it is a definitional fact of the term. What may be – and is – questioned is when a certain BMI is indeed unhealthy. That is, the evaluative aspects seem to have a secondary status as projections on the fundamentally descriptive or natural concept of BMI. The descriptive part, however, is rock solid.

The normative aspects argued for in the previous subsection, however, are hard to picture as secondary projections on the descriptive aspects. The point of the critique of a natural, non-normal conception of risk has been that normative aspects are prevalent in the very *ascription* of the quantitative risk, and not only in the evaluation of a previously already given

quantification. Hence, in order to *establish* the magnitude of the harm, or the way in which the probability and severity are to be combined into a risk measure, we make evaluations that are over-and-above the natural basis of these notions.

For such notions where there is no clear primacy of a technical definition, philosophers in the thick concept debate have been increasingly skeptical of the reductive idea. Philosophers such as John McDowell, Bernard Williams, and Jonathan Dancy have argued that no such separation of a descriptive, autonomous part on the one side and an evaluation on the other can be given (McDowell 1978, 1979, 1981; Williams 1985; Dancy 1995, 2004). The only way to understand a thick concept is to understand the descriptive and evaluative aspects *as a whole*. The idea is that for a thick concept, the evaluative aspect is profoundly involved in the practice of using it, and therefore one cannot understand a thick concept without understanding also its evaluative point. (I here use the epistemological framing of the claim used by McDowell. Dancy (1995) goes even further, making the metaphysical claim that the descriptive and evaluative content is an indissoluble amalgam.)

On the analysis of risk and safety given in this chapter, it is hard to see the normativity of the aspects of risk and safety as a secondary, projective aspect of a notion that is otherwise essentially descriptive. The main reason for this is that the normative aspects of risk figured not only as an evaluation of the “finished product,” as in the BMI example and in the question of the acceptability of the risk, but on the very ascription of the core aspects of risk, such as harm and epistemic uncertainty, and the delimitation and distributive aspects.

For risk and safety, thus, there is an essential interdependence between the natural-descriptive aspects and the normative-evaluative aspects. That an event has a higher expected value of death than another event – treated as a natural input – is typically a reason to ascribe the former a higher risk than the latter. But this is not the case if there are other harms than death that should count as well, or a questionable distribution of the harmful outcomes for the latter case (as in the staff in the novel nuclear plant), or if the epistemic security is much higher in the latter case, etc. – all arguably normative aspects. In this respect, risk and safety are unlike paradigmatic natural concepts such as gold. Even if people treated gold as something positive, as many do, this does not change the fact that only substances of a certain atomic number counts as gold: descriptive notions are all that matters. And it is unlike concepts such as BMI, or arguably even concepts as *fat* – on its quantitative notion at least – where descriptive aspects are all that counts for identifying something as having a higher BMI, or as being fatter.

On this reading, risk and safety seem to be of the same essential thick kind that philosophers such as John McDowell, Bernard Williams, and Jonathan Dancy argue cannot be separated into a natural part and a normative part, but must be treated as concepts that are *both* natural and normative, and for which no natural reduction can be given that suffices to pick out the extension of the terms. Currently, this anti-reductive camp is the dominant one in the debate about thick concepts, and also theorists such as Simon Blackburn who initially defended the reductive view (Blackburn 1984) have later abandoned it (Blackburn 1998). Even if Hare defended a reductivist strategy as late as in his 1997: 61, contemporary separatists argue instead for a weaker claim, namely, that a thick concept can be separated into a thin normative concept and a description, but that for the extension of a thick concept, the normative part is needed as well (Elstein and Hurka 2009). In other words, the output of recent moral philosophy is skepticism of the reductive claim for thick concepts such as risk and safety.

## Risk, Objectivity, and Social Constructions

Let us now assume, in line with the arguments in the previous subsection, that risk and safety are thick normative concepts that cannot be reduced to natural ones. Is this bad news for proponents of the scientist or psychological approach to risk? Many theorists in the risk debate seem to make a close connection, if not equivalence, between a notion's status as natural or descriptive, and its being an objective concept. If a concept is not a natural concept, they seem to assume it is subjective rather than objective. It is important to note, however, that this fails to appreciate the large debate in philosophy concerning the status of the field of the normative, in particular the moral domain. In fact, many of the main proponents of the irreducible normativity of thick concepts mentioned above, for example, McDowell and Dancy, are moral realists who believe that there are objective truths in moral discourse. Hence, it is important to notice that claiming that risk is normative is not the same thing as claiming that it is non-objective.

In this subsection, I will investigate a common critique of the notions of risk and safety that seem to claim something stronger than mere normativity, namely, that these notions are essentially subjective or socially constructed. Objectivity and reality are philosophical questions that run deep very fast, and it would take us too far to argue for any particular position in this chapter. The current aim is rather to clarify the extent to which claims of subjectivity of risk, or even the non-reality of risk, are strong and radical claims.

Many theorists from the social sciences that approach risk on what I have called the *cultural perspective* argue that risk is a social construct, which is often labeled as a subjective, social feature rather than a feature of objective reality. People from different cultures as well as within one culture have very different views on what constitutes a risk and how severe a risk may be, they point out, and they go on to claim that there is no fact of the matter over and above these individual or cultural views (Douglas and Wildavsky 1982; Wynne 2000; Slovic 2000). Judith Bradbury, for example, describes the cultural view as the view that risk is conceived as "a socially constructed attribute, rather than as a physical entity that exists independently of the humans who assess and experience its effects" (Bradbury 1989, p. 381).

It should be noted that subjectivity in Bradbury's sense is a rather radical view. It means that the risk of climbing the ascent of K2, or driving 150 km/h in New York City, depends on humans assessing these activities rather than on physical facts that exist independently of such activities. In order to see the radicalism in such a position, and question whether the basic insights of the social constructivist idea motivate such radicalism, let us distinguish between three senses in which something can be socially constructed.

Let us call the first sense of social construct (essential) *reference-dependence*. A social system is needed in order for such an entity to exist. The paradigmatic example here is money (Searle 1995): in order for a piece of paper, metal object, or other physical entity to be money, it has to play a certain role in a social system. Before humans invented the concept of money and established the system in which it had a place, there was no such thing as money. If we take away the system, if we moved away from that sort of economy, or if intelligent life ceased to exist, there would be no money anymore.

The second sense of social construct is weaker: *interactive-dependence* (Hacking 1999). An entity that is interactive-dependent is not essentially dependent on any social system in which it is conceptualized. Still, its conceptualization in a social context affects the entity thus categorized. *Woman*, for example, is such an interactive category. By effectively categorizing someone

as a woman – or stop doing so – one is affecting the way in which the person is viewed by others and herself, which behavior is typical and expected, right and wrong. Before we had a *concept* of woman, there existed women, in the sense that the kind of entities existed that (sufficiently) fit the concept. Its introduction, however, has changed the very category: a basic insight of feminism is that categorizing someone as a woman affects how she conceives of herself, which is one of the important problems for societal change. It is why merely *formally* allowing women entrance to societal positions and roles that used to be restricted to men does not by itself establish a substantial shift in the social structure, even if there are no essential biological restrictions that prohibit such a shift.

The third sense of social construct is the weakest: *sense-dependence*. (The distinction between reference-dependence and sense-dependence is taken from Brandom (2002, pp. 194–195), although it is here used in a slightly different way.) The idea here is that the entity in question is neither interactive-dependent nor essentially reference-dependent on its conceptualization and the social system in which it figures. However, our being able to *grasp* the concept and thus classify objects and properties in the world in accordance with it is dependent on a social context in which it has evolved. An example is the concept of a quark. The concept of a quark is dependent on the modern development of physics. While the concept of a quark was developed independently by Murray Gell-Mann and George Zweig in 1964, their being able to develop it was of course the fruit of social arrangements – ideas and institutions – throughout centuries. Indeed, all complex concepts such as those of theoretical entities in natural science theories are the fruit of social construction in this sense.

This third sense of the socially constructed is by far the weakest. Most philosophers of language would admit that *any* concept is socially constructed in the sense of having been developed in a social environment, for the simple reason that language *in total* is a social construct. Now, the extent of that sociality, and how necessary – as opposed to merely historically contingent – it is for language to be thus developed, may be disputed, but it seems to be an undeniable fact that the ability to express claimable content to other beings, to categorize objects and events in order to refer to them by language, is a social affair. How can it not be? Still, there is certainly a difference between the level of social exchange that is needed for different concepts. The concept of quarks need a bit more social lever – the development of advanced physics – in order to be constructed than, for example, the concept of food, which is closely connected to basic human needs.

Equipped with these three senses of social constructs, the one that most clearly fits Bradbury's description of risk as a "socially constructed attribute, rather than as a physical entity that exists independently of the humans who assess and experience its effects" is the first one, the essential reference-dependence. In the two other senses it seems (and I have assumed in the explanation of them) that while the *concept* is a social construct, the *entity* may exist independent of the humans that assess and experience its effects. It does make sense to state that women – and men, and children, and humans in general – existed prior to the construction of a concept to sort them out as such. Or, at the very least, further argument is needed in order to show why this is not so. For reference-dependent concepts such as money, it is the other way around: it makes little sense to claim that money existed before the conceptual practice involving it was in place.

In the debate, the range of claims sorted under the banner of social constructivism is very broad. Often in practice the core claim seems to be something closest to the third sense of social construct, which primarily focuses not on the entity as such but on the *concept* we have

developed. For example, take Summerton and Berner's description of the main premise of the cultural approach:

- What is identified as a risk is not based on an *a priori* "fact" as identified by technical or medical experts; neither are risks merely rule-of-thumb assessments made by isolated individuals. The ways in which individuals – including experts – interpret risk can instead be seen as an expression of socially located beliefs and world views that to a large extent stems from the individual's situated position and experiences within social hierarchies, institutions and groups (Summerton and Berner 2003, pp. 6–7).

Their claim is much less radical than Bradbury's, since they explicitly limit themselves to what is *identified* as a risk by individuals, claiming first that it is not the same as what experts take it to be, and second that it is, for all individuals including experts, to a large extent a consequence of her social environment. None of this is a denial of the possibility of risks *existing* independently of individuals experiencing them.

Indeed, some theorists make clear that they are not arguing against objectivity per se. Clarke and Short point out that "objects might be dangerous in some objective sense" (Clarke and Short 1993, p. 79). The basic task on the cultural view, however, "is to explain how social agents create and use boundaries to demarcate that which is dangerous" (Clarke and Short 1993). In other words, the motive for making a claim of social construction in a domain can be to underline an explanatory focus: that of investigating how social circumstances, such as organizational environments, social hierarchies or other aspects of our socio-cultural context, play a part in how we conceptualize our world as well as what we take to be the extension of those concepts.

This explanatory task may of course be of paramount importance for how we should arrange our investigations into risks, and how to make decisions about them. Furthermore, it may convincingly show how our conceptions feed back into the phenomenon itself, that is, give detailed insights into the interactive-dependence of risk and safety. (Note that risk and safety are only indirectly interactive-dependent, since *we* are the ones affected by something's being categorized as risky – thus possibly changing the risk at hand – not the risk-entity itself.)

Importantly, however, neither sense-dependence nor interactive-dependence entails subjectivity. There may still perfectly well be an objective fact of the matter whether it is safe to climb K2, even if the actual safety depends on how we have perceived of it, as mediated, for example, by its manifestations in our training and preparation before attempting it, or in how we react to things that happen while climbing. Hence, even if risk and safety are both sense-dependent and interactive-dependent concepts, they can still be objective concepts corresponding to real properties in the world.

There are, however, claims that are made under the heading of social constructivism that deny the reality of risk (e.g., Ewald 1991, p. 199; Dean 1999, p. 177). Also a risk psychologist such as Paul Slovic makes this inference from constructivism to non-realism in a rather recent paper when he contrasts danger with risk, claiming that "danger is real, but risk is socially constructed." (Slovic 1999, p. 689) It is hard, though, to see the motivation for this denial of reality status for risk. Slovic's justification is representative: "Risk assessment is inherently subjective and represents a blending of science and judgment with important psychological, social, cultural, and political factors" (Slovic 1999). But even if we grant that risk *assessment* is subjective – for example, that it has uncorrected biases from cultural and political factors that result in faulty norms – this does not mean that risks cannot be real. Note that even if the *assessment* of our oil

reserves are biased by wishful thinking as well as political and economic agendas, in addition to pure scientific uncertainty in the current models of estimation, it does not follow that oil is not real (or that Eve can have less oil in the tank of her car than Jones has in his).

If no further arguments are presented, it seems that the claims of subjectivity and anti-realism are unjustified inferences from sound insights into the societal nature of concept grasping and assessments of phenomena. In Slovic's claim in particular, this impression is enhanced from the seemingly arbitrary demarcation line between risk and danger. What constitutes a danger (or hazard) seems to be just as connected to human affairs as risk. So if danger is real, why not risk?

Lastly, we should note that even the strongest version of dependence I have mentioned, essential reference-dependence, in itself entails neither subjectivity nor anti-realism. Take the concept of money. The common wisdom is that money is indeed real, and claims concerning them may be objectively true or false, even if it is also true that the existence of potential money facts such as "My mobile phone is now worth roughly \$200 less than when I bought it" or "I sold my car for \$2,000" is dependent on a complex mix of social arrangements.

Naturally, none of the distinctions made in this subsection are by themselves an *argument* for the objectivity or reality of risk. Rather, the aim has been to distinguish between some of the different types of claims that the overall label of social construction may amount to, and to make plausible that denial of objectivity of risk and – even more – the reality of risk lies at the far end of this spectrum.

## Further Research

---

As emphasized at the outset, risk and safety research is a heterogeneous field, and there are many interesting areas for research.

## Analysis of Key Concepts

---

*Mapping different risk and safety conceptions.* In this chapter, several common interpretations of risk and safety have been discussed. Although it is well known that different people use the key concepts of risk research differently, a better mapping of the differences in usage of "risk" and "safety" in various fields is much needed. In particular, this is important for the psychological research where people's attitudes toward risks are the target, and where the terms are generally less clearly defined than in technical areas.

*Risk versus safety.* As indicated in this chapter, risk and safety are frequently treated as antonyms. While this may often be harmless, this treatment is not always viable. For example, it hides a difference in focus between risk research and safety research that should be more clearly explored.

*Uncertainty.* Uncertainty is an aspect of (lack of) safety that is sometimes hidden by the focus on handling the risk. Epistemic uncertainty is an important aspect of safety ascriptions, but, as emphasized in this chapter, there is large controversy as to how this notion is to be measured. Moreover, in many areas of risk and safety, the traditional method of coping with uncertainty is by an increased emphasis on probabilistic safety analysis (PSA). Sometimes, as in for example the nuclear industry, this has amounted to a very high level of descriptive sophistication. Even on this sophisticated level of PSA description, however, there are residual

uncertainties, uncertainties that cannot be captured by PSA alone. Finding ways of handling these remaining uncertainties is an important task.

*Safety versus security.* While the concept of security is strongly interrelated with the concept of safety, there are important differences in the actual *practices* of the two fields. For example, security measures are more focused on protection against intentional threats, whereas safety measures have been directed against non-intentional harmful events as well as utilizing, to a greater extent, probabilistic methods. Investigating the differences and similarities in how safety and security are understood can be an important way to develop our understanding of both notions, and, consequently, develop our methods of ensuring safety and security.

*Safety culture.* The notion of a safety culture is studied by the social and behavioral sciences and employed by safety professionals. But it is rather unclear what a safety culture consists in. Is it an evaluative concept, and what counts as a good or bad safety culture? Can an organization lack a safety culture, or is that equal to a bad one? And what does “culture” signify in the context? Since it is considered an important “tacit” aspect in reaching safety, more conceptual clarity about the notion is called for.

*The status of risk and safety.* In this chapter, we touched on the complicated issues about the status of risk and safety. In addition to the overarching questions of naturalness, normativity, and constructivism, there are also interesting questions of subjectivity and objectivity when it comes to certain aspects of risk and safety. For harm and probability in particular, there are both subjective and objective interpretations. How these interpretations are best understood, and how this matters for risk and safety research, are pressing interdisciplinary concerns.

## Further Issues

---

Related to these conceptual investigations, there are methodological, epistemic, and ethical issues for future research. Examples of such topics include:

*Knowledge: implicit and explicit.* Although effort is made in articulating explicit, propositional accounts of risk and safety, a significant amount of knowledge of risk and safety remains tacit in an organization. In connection with concepts such as uncertainty and safety culture, the relation between tacit and explicit knowledge of risk and safety should be investigated.

*The sequential process approach.* The viability of the common sequential process in safety analysis of starting with a scientific assessment of the risk, followed by an evaluative step in which the overall safety decision is made, depends on the viability of suitably isolating the descriptive, scientific aspects of the risk at hand. Any limitations to this approach have methodological consequences, not only for how we should go about gaining acceptance for a certain risk, or how we should compare risks to the advantages of the technology causing the potential risk, but for the actual enterprise of gaining *knowledge* of the risks involved. A common assumption in risk analysis is that ethical aspects come as merely an *addition* to the risk assessment, something to be considered in risk management only – an assumption that may be questioned, as in the arguments in this chapter that evaluative aspects surfaces already “inside” the very assessment of the risk. An important area of future research involves investigating the consequences of the internal as well as external presence of ethical aspects in risk assessment and management.

*New versus old.* Trying out new methods and techniques is paramount for technological progress. Yet, this means by definition less tested in actual circumstances, which often implies

a large epistemic uncertainty. The circumstances when it is ethically correct to use new designs in potentially dangerous facilities is an interesting question in need of further research, both in general and in the many applied areas of risk and safety research.

*Control versus integrity.* Another trade-off of interest is the one between on the one hand surveillance and detailed supervision or control, and on the other, integrity and autonomy at the workplace. Here, what is deemed most safe and what is deemed as respecting the right to integrity and autonomy may not go hand in hand, and what to do with cases when these ideals clash is a central moral question in risk and safety ethics.

*Disagreement.* As indicated throughout this chapter, there are many different conceptions of safety, as well as many different ways of assessing the very same conception (such as the expected value conception). How should we best treat disagreement about risk and safety? Some disagreements are about scientific issues proper, and these seem reasonable to solve within the scientific community. Also scientists disagree, however, and sometimes on a level that is significantly relevant for the risk assessment. What should we do in such circumstances? But even when there is no disagreement of the underlying natural properties, we may disagree about the normative aspects of risk and safety, as well as about further ethical considerations. How we best treat such different cases of disagreement is a pressing topic for future study.

## Conclusion

---

In this chapter, I have analyzed several aspects of the concepts of risk and safety. As I started by pointing out, there are many different – if overlapping – perspectives from which the questions of risk and safety are treated, as well as several interpretations of the terms “risk” and “safety” that are relevant in the context of societal decision-making. An evident, but still often neglected consequence of this heterogeneous set of interpretations is that we should always keep in mind the possibility of talking past each other in matters involving risk and safety. There are many viable ways of using the terms, and ascertaining that we are on the same track is a minimal condition for risk and safety communication.

The main focus in the chapter has then been on the quantitative notions of risk and safety. Central aspects such as probability, harm, and uncertainty have been treated. In particular, the controversial question of how risk and safety should be inferred from these aspects – and possibly even additional ones – should be kept in mind. Moreover, whether risk and safety are natural concepts or contain normative aspects that cannot be reduced to natural ones are important questions for how we should understand these concepts. Finally, I pointed out that even if risk and safety are normative concepts – as I indeed believe they are – this is not equal to the claim that risk and safety cannot be objective or that they are constructed in any deep sense. Still, risk and safety are central concepts in societal decision-making, and both scientific and extra-scientific concerns are vital if we are to make the best decisions when risks are involved.

## Acknowledgments

---

The research leading to these results has received funding from the European Community’s Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 237590.

## References

- Asveld L, Roeser S (eds) (2009) *The ethics of technological risk*. Earthscan, London
- Aven T (2003) Foundations of risk analysis: a knowledge and decision-oriented perspective. Wiley, Chichester
- Aven T (2007) A unified framework for risk and vulnerability analysis and management covering both safety and security. *Reliab Eng Syst Safe* 92:745–754
- Aven T, Renn O (2009) On risk defined as an event where the outcome is uncertain. *J Risk Res* 12:1–11
- Baron J (1987) Second-order probabilities and belief functions. *Theory Decision* 23:25–36
- Blackburn S (1984) Spreading the word. Clarendon, Oxford
- Blackburn S (1998) *Ruling passions: a theory of practical reasoning*. Clarendon, Oxford
- Boyd R (1991) Realism, anti-foundationalism and the enthusiasm for natural kinds. *Phil Stud* 61: 127–148
- Bradbury J (1989) The policy implications of differing concepts of risk. *Sci Technol Hum Values* 14: 380–399
- Brandom R (2002) *Tales of the mighty dead: historical essays in the metaphysics of intentionality*. Harvard University Press, Cambridge, MA
- Campbell S (2005) Determining overall risk. *J Risk Res* 8:569–581
- Clarke L, Short J (1993) Social organization and risk: some current controversies. *Annu Rev Sociol* 19: 375–399
- Cohen BL (2003) Probabilistic risk analysis for a high-level radioactive waste repository. *Risk Anal* 23: 909–915
- Dancy J (1995) In: Defence of thick concepts, in French, Uehling, Wettstein (eds), *Midwest studies in philosophy* 20. University of Notre Dame Press, Notre Dame, Ind
- Dancy J (2004) *Ethics without principles*. Clarendon, Oxford
- de Finetti B (1937) La prévision: see lois logiques, ses sources subjectives, *Annales de l'Institut Henri Poincaré* 7
- Dean M (1999) *Governmentality: power and rule in modern society*. Sage, London
- Douglas EJ (1983) *Managerial economics: theory, practice and problems*, 2nd edn. Prentice Hall, Englewood Cliffs
- Douglas M, Wildavsky A (1982) *Risk and culture: an essay on the selection of technological and environmental dangers*. University of California Press, Berkeley
- Dubois D, Prade H (1988) Decision evaluation methods under uncertainty and imprecision. In: Kacprzyk J, Fedrizzi M (eds) *Combining fuzzy impression with probabilistic uncertainty in decision making*. Springer, Berlin, pp 48–65
- EC (2003) Technical guidance document in support of commission directive 93/67/EEC on Risk Assessment for new notified substances, Commission Regulation (EC) No 1488/94 on Risk Assessment for existing substances and Directive 98/8/EC of the European Parliament and of the Council concerning the placing of biocidal products on the market. Joint Research Centre, EUR 20418 EN, Office for Official Publications of the EC, Luxembourg
- Ellsberg D (1961) Risk, ambiguity, and the savage axioms. *Quart J Econ* 75:643–669
- Elstein D, Hurka T (2009) From thick to thin: two moral reduction plans. *Canad J Philos* 39:515–536
- Ewald F (1991) Insurance and risk. In: Burchell G, Gordon C, Miller P (eds) *The Foucault effect: studies in governmental rationality*. Harvester Wheatsheaf, Hemel Hempstead
- Fischhoff B, Slovic P, Lichtenstein S, Read S, Combs B (1978/2000) How safe is safe enough? A psychometric study of attitudes toward technological risk and benefits. In: Slovic P (ed) *The perception of risk*. Earthscan, London, pp 80–103
- Gärdenfors P, Sahlin N-E (1982/1988) Unreliable probabilities, risk taking, and decision making. In: Gärdenfors P, Sahlin N-E (eds) *Decision, probability, and utility*. Cambridge University Press, Cambridge, pp 313–334
- Graham JD, Weiner JB (eds) (1995) *Risk versus risk: tradeoffs in protecting health and the environment*. Harvard University Press, Cambridge, MA
- Hacking I (1991) A tradition of natural kinds. *Phil Stud* 61:109–126
- Hacking I (1999) *The social construction of what?* Harvard University Press, Cambridge, MA
- Hansson SO (1998) *Setting the limit. Occupational health standards and the limits of science*. Oxford University Press, Oxford
- Hansson SO (2004a) Philosophical perspectives on risk. *Technie* 8:1
- Hansson SO (2004b) Weighing risks and benefits. *Topoi* 23:145–152
- Hansson SO (2005) Seven myths of risk. *Risk Manage* 7:7–17
- Hansson SO (2009) Risk and safety in technology. In: Meijers A (ed) *Handbook of the philosophy of science*, vol 9: *philosophy of technology and*

- engineering sciences. Elsevier, Amsterdam, pp 1069–1102
- Hansson SO (2010) Risk: objective or subjective, facts or values. *J Risk Res* 13:231–238
- Hare RM (1952) The language of morals. Clarendon, Oxford
- Hare RM (1997) Sorting out ethics. Clarendon, Oxford
- Howson C, Urbach P (2006) Scientific reasoning: the Bayesian approach, 3rd edn. Open Court Publishing Company, Chicago
- IAEA (2000) Safety of nuclear power plants: design. International Atomic Energy Agency, Vienna
- ISO (2002) Risk management – vocabulary – guidelines for use in standards. ISO/IEC Guide 73:2002
- Kaplan M (1983) Decision theory as philosophy. *Phil Sci* 50:549–577
- Knight FH (1921) Risk, uncertainty and profit. BoardBooks, Washington, DC (Reprinted 2002)
- Kraus N, Malmfors T, Slovic P (1992/2000) Intuitive toxicology: experts and lay judgements of chemical risks. In: Slovic P (ed) *The perception of risk*. Earthscan, London, pp 285–315
- Kuhn T (1962) *The structure of scientific revolutions*. University of Chicago Press, Chicago
- Kyburg HE (1968) Bets and beliefs. *Am Phil Q* 5:54–63
- Lakatos I, Musgrave A (eds) (1970) *Criticism and the growth of knowledge*. Cambridge University Press, London
- Lange M (ed) (2007) *Philosophy of science: an anthology*. Blackwell, Malden
- Le Grand J (1991) *Equity and choice: an essay in economics and applied philosophy*. Harper Collins Academic, London
- Levi I (1974) On indeterminate probabilities. *J Phil* 71:391–418
- Levi I (1986) Hard choices: decision making under unresolved conflict. Cambridge University Press, Cambridge
- Lewens T (2007) *Risk: philosophical perspectives*. Routledge, New York
- Little M (1994) Moral realism II: non-naturalism. *Phil Books* 35:225–232
- Lowrance W (1976) Of acceptable risk – science and the determination of safety. William Kaufmann, Los Altos
- Mayo D (1991) Sociological versus metascientific views of risk assessment. In: Mayo D, Hollander R (eds) *Acceptable evidence, science and values in risk management*. Oxford University Press, Oxford, pp 249–280
- McDowell J (1978) Are moral requirements hypothetical imperatives? In: *Proceedings of the Aristotelian society*, Supplementary volume 52, pp 13–29
- McDowell J (1979) Virtue and reason. *Monist* 62:331–350
- McDowell J (1981) Non-cognitivism and rule-following. In: Holtzman S, Leich C (eds) *Wittgenstein: to follow a rule*. Routledge/Kegan Paul, London/Boston, pp 141–162
- McMullin E (1982) Values in science. In: *PSA: proceedings of the biennial meeting of the philosophy of science association*, 2:3–28
- Miller CO (1988) System safety. In: Wiener EL, Nagel DC (eds) *Human factors in aviation*. Academic, San Diego, pp 53–80
- Misumi Y, Sato Y (1999) Estimation of average hazardous-event-frequency for allocation of safety-integrity levels. *Reliab Eng Syst Safe* 66:135–144
- Möller N (2009a) Thick concepts in practice: normative aspects of risk and safety. Royal Institute of Technology, Stockholm
- Möller N (2009b) Should we follow the experts' advice? Epistemic uncertainty, consequence dominance and the knowledge asymmetry of safety. *Int J Risk Assess Manage* 11:219–236
- Möller N (2010) The non-reductivity of normativity in risks. In: de Vries MJ, Hansson SO, Meijers AWM (eds) *Norms and the artificial: moral and non-moral norms in technology* (Forthcoming)
- Möller N, Hansson SO (2008) Principles of engineering safety: risk and uncertainty reduction. *Reliab Eng Syst Safe* 93:776–783
- Möller N, Hansson SO, Peterson M (2006) Safety is more than the antonym of risk. *J Appl Philos* 23:419–432
- Moore GE (1903) *Principia ethica*. Cambridge University Press, Cambridge
- National Research Council (1983) *Risk assessment in the federal government managing the process*. National Academy Press, Washington, DC
- Otway H (1987) Experts, risk communication, and democracy. *Risk Anal* 7:125–129
- Peterson M (2002) What is a de minimis risk? *Risk Manage* 4:47–55
- Ramsey F (1931) Truth and probability. In: Braithwaite RB (ed) *The foundations of mathematics and other logical essays*. Routledge & Kegan Paul, London, pp 156–198. (Reprinted in: Gärdenfors P, Sahlin N-E (eds). *Decision, probability, and utility*. Cambridge University Press, Cambridge, pp 19–47)
- Rechard RP (1999) Historical relationship between performance assessment for radioactive waste disposal and other types of risk assessment. *Risk Anal* 19:763–807
- Renn O (1992) Social theories of risk. In: Krimsky S, Golding D (eds) *Social theories of risk*. Praeger, Westport
- Renn O (2005) Risk governance: towards an integrative approach. White paper no. 1, written by Ortwin Renn with an Annex by Peter Graham. International Risk Governance Council, Geneva

- Resnik M (1987) Choices: an introduction to decision theory. University of Minnesota Press, Minneapolis
- Rosa EA (1998) Metatheoretical foundations for post-normal risk. *J Risk Res* 1:15–44
- Savage L (1972/1954) The foundations of statistics, 2nd edn. Dover, New York
- Searle J (1995) The construction of social reality. The Free Press, New York
- Sen A (1987) On ethics and economics. Blackwell, Oxford
- Shrader-Frechette K (1993) Burying uncertainty: risk and the case against geological disposal of nuclear waste. University of California Press, Berkeley
- Skyrms B (1980) Higher order degrees of belief. In: Mellor DH (ed) Prospects for pragmatism. Cambridge University Press, Cambridge, pp 109–137
- Slovic P (1999) Trust, emotion, sex, politics, and science: surveying the risk-assessment battlefield. *Risk Anal* 19:689–701
- Slovic P (2000) The perception of risk. Earthscan, London
- Soames S (2005) Philosophical analysis in the twentieth century: the age of meaning. Princeton University Press, Princeton
- Stevenson C (1944) Ethics and language. Yale University Press, New Haven
- Summerton J, Berner B (2003) Constructing risk and safety in technological practice: an introduction. In: Summerton J, Berner B (eds) Constructing risk and safety in technological practice. Routledge, London, pp 1–19
- Tench W (1985) Safety is no accident. Collins, London
- Unwin S (1986) A fuzzy set theoretic foundation for vagueness in uncertainty analysis. *Risk Anal* 6:27–34
- Vallentyne P (1998) The nomic role account of carving reality at the joints. *Synthese* 115:171–198
- von Neumann J, Morgenstern O (1944) Theory of games and economic behavior. Princeton University Press, Princeton
- Wandall B (2004) Values in science and risk assessment. *Toxicol Lett* 152:265–272
- Williams B (1985) Ethics and the limits of philosophy. Harvard University Press, Cambridge, MA
- Willis HH (2007) Guiding resource allocations based on terrorism risk. *Risk Anal* 27:597–606
- Wolf M (2002) The curious role of natural kind terms. *Pacific Phil Q* 83:81–101
- Wynne B (1982) Institutional mythologies and dual societies in the management of risk. In: Kunreuther H, Ley E (eds) The risk analysis controversy. Springer, Berlin, pp 127–143
- Wynne B (1992) Carving out science (and politics) in the regulatory jungle. *Soc Stud Sci* 22:745–758



# 4 Levels of Uncertainty

Hauke Riesch

University of Cambridge, Cambridge, UK

<i>Introduction</i> .....	88
<i>Background</i> .....	89
<i>Objects of Uncertainty</i> .....	96
Examples .....	101
The Lottery .....	101
Saving Our Bacon .....	102
Carbon Capture and Storage .....	105
Climate Change .....	107
<i>Further Research</i> .....	108

**Abstract:** There exist a variety of different understandings, definitions, and classifications of risk, which can make the resulting landscape of academic literature on the topic seem somewhat disjointed and often confusing. In this chapter, I will introduce a map on how to think about risks, and in particular uncertainty, which is arranged along the different questions of what the different academic disciplines find interesting about risk. This aims to give a more integrated idea of where the different literatures intersect and thus provide some order in our understanding of what risk is and what is interesting about it. One particular dimension will be presented in more detail, answering the question of what exactly we are uncertain about and distinguishing between five different levels of uncertainty. I will argue, through some concrete examples, that concentrating on the objects of uncertainty can give us an appreciation on how different perspectives on a given risk scenario are formed and will use the more general map to show how this perspective intersects with other classifications and analyses of risk.

- ▶ I beseech you, in the bowels of Christ, think it possible you may be mistaken (Oliver Cromwell, addressing the Church of Scotland, 1650) (From Carlyle 1871).

## Introduction

---

What we mean by risk is not a clear issue because many writers use the word with slightly different meanings and definitions, even beyond the more vague everyday usage of the term. Aven and Renn (2009), for example, have found 10 different definitions they gathered from the wider risk literature. The problem of clear terminology continues if we go into the various classifications and clarifications of risk and uncertainty, with scholars distinguishing between risks, uncertainties, indeterminacies, ambiguities, and levels, objects or locations of risk and/or uncertainty. With this in mind, I feel slightly apologetic about writing about another scheme devised by myself and David Spiegelhalter (Spiegelhalter 2010; Spiegelhalter and Riesch *in press*), where we use, again, our own terminology, this time in trying to distinguish between different things we can be uncertain about. In this chapter, I will try to explain our distinctions and where they correlate and/or fit in with other classifications of risk and uncertainty, as well as provide an argument on why we feel this particular classification adds to the literature on risk theory by going through a couple of real-world examples.

As Norton et al. (2006) note in their reply to the paper by Walker et al. (2003) discussed below, “an important barrier to achieving a common understanding or interdisciplinary framework is the diversity of meanings associated with terms such as ‘uncertainty’ and ‘ignorance,’ both within and between disciplines” (Norton et al. 2006, p. 84). The proliferation of what we mean by risk and how we categorize it within the literature is partly due to the different agendas the different disciplines have with regard to the topic. The question “what do we want to know about risk?” will be answered differently by scholars, for example, interested in risk perception and those interested in the “risk society.” Asking this question explicitly may help us in finding out where the different disciplinary approaches to risk intersect. Our classification is partly intended to do just that, mostly because I (as a sociologist) and David

Spiegelhalter (as a statistician) have always had slightly different conceptions of what is academically interesting about risk, and our collaboration was partly an attempt to build a conception of risk which is useful for both social and scientific/technical disciplines and will be useful for communicating across this divide by giving a clear account of how and why, in Funtowicz and Ravetz's phrase, there is "a plurality of legitimate perspectives" on risk (Funtowicz and Ravetz 1993 p. 739). At the same time, I hope it will provide a useful and relatively simple map through which the different academic disciplines' interests in risk can be compared and connections seen more easily.

In this chapter, I want to advance the idea that one can confront the different meanings which risk is given and offer an idea of how they are related. There exist many other schemes that try to categorize risks such as Renn and Klinke (2004), Stirling (2007), or van Asselt and Rotmans (2002), and I will try to show how they fit in to our overall picture in the following section. As a departing point, I take risk to mean roughly a function of the uncertainty of an outcome and its impact. This definition leaves room for plenty of uncertainties itself, especially since there is no agreement on how to measure impact, or how to compare impacts of completely different categorization, and there is plenty of literature in risk studies devoted to this problem.

The uncertainty part of risk however is itself very problematic. There are some uncertainties we can put a number on, some where we can only evaluate qualitatively and some we have absolutely no idea on how to even start evaluating them. The classification I will propose here is meant to bring some order into the way we think about uncertainty and provide a way in which different types of uncertainty and its classifications can be, if not directly compared, at least brought under the same scheme. Comprehensive surveys of what uncertainties and risks really are and how they should be classified can easily lead to a rather complicated structure that becomes less useful as a heuristic tool for people working within risk. This is more so on the social, policy, and communication aspects than in the technical risk assessment areas, for whom such schemes will be more useful, and I will concentrate on the former in this chapter. Out of the many different dimensions in which uncertainty can be categorized, we chose one in particular which we believe is most helpful when we seek to understand how different people and groups conceptualize and react toward risks. It is meant to analyze risks according to the following question: What kind of thing exactly are we uncertain about?

## Background

---

Philosophical classifications of probability have traditionally focused on questioning where our uncertainty derives from, with the two main choices being uncertainty inherent in the system, and uncertainty arising from our incomplete knowledge. These two interpretations of probability are named by philosophers (Hacking 1975, see also Gillies 2000) *epistemic* probability and *aleatoric* (also often called *ontological* or *ontic*) probability. This basic distinction still underlies modern philosophical theories of probability and can be seen, for example, in the philosophical split between Bayesian (subjective) and frequentist (objective) interpretations of probability in statistics (see also Gillies 2000).

Uncertainty in a larger sense, as opposed to the mathematically defined concept of probability, has also seen attempts at classification. An early and very influential distinction

came from Frank Knight, who distinguished uncertainties which are quantifiable which he called risks, and those that are not quantifiable, which he called uncertainties:

- The essential fact is that 'risk' means in some cases a quantity susceptible of measurement, while at other times it is something distinctly not of this character; and there are far-reaching and crucial differences in the bearings of the phenomena depending on which of the two is really present and operating. [ . . . ] It will appear that a measurable uncertainty, or 'risk' proper, as we shall use the term, is so far different from an unmeasurable one that it is not in effect an uncertainty at all (Knight 1971 [1921]).

This classification has proved to be very influential especially among sociologists, but is in my opinion slightly unfortunate as it propagates confusion with the traditionally defined concept of risk equaling probability times outcome (or, in the more modern sense focusing on negative outcomes, probability times harm). Although I recognize the usefulness of Knight's distinction for this work, to avoid confusion I prefer to work with the conception that risk refers to a measure of uncertainty combined with the potential outcome.

Combining these two perspectives in a sense, Stirling (2007) recently proposed to divide both the uncertainty as well as the outcome aspects of risk into "problematic" versus "unproblematic" in a similar way to which Knight distinguished between quantifiable and unquantifiable uncertainty. This results in a two by two matrix: at the corner where the probabilities as well as (our knowledge of) the outcomes are unproblematic there are risks associated with the typical statistical risk analyses such as Monte Carlo simulations or cost-benefit analyses – these scenarios he terms "risks" in the traditional sense used by most scientists and risk analysts. Scenarios where the probability is knowable, but we are more unsure about the outcomes, he terms "ambiguities"; risks where conversely the outcomes are unproblematic but the probabilities are, he calls "uncertainties." When neither are unproblematic, he talks about conditions of "ignorance." It is worth also pointing out that the term "ambiguity" is used in other disciplines, for example, behavioral economics, to mean unknown probabilities, which is almost precisely the opposite to Stirling's sense – this demonstrates, again, the problems of terminology within the wider risk literature. Technology assessment on the other hand traditionally uses similar terminology but without taking Stirling's ambiguity into account.

Stirling argues that dividing risks into these categories can give us guidance on the circumstances when the precautionary principle could be a valid rule: by dividing risks into qualitatively distinct groups, he argues that the principle can be an important rule for helping with decision making in those circumstances where the outcomes or probabilities are not well understood, and no other type of decision rule would otherwise be helpful.

Another influential attempt at classifying risk elaborated to inform risk assessment policy eventually evolved to inform Funtowicz and Ravetz's very influential concept of postnormal science (Funtowicz and Ravetz 1990, 1993). Funtowicz and Ravetz proposed to map risks as a measure of uncertainty and impact ("decision stakes") and claimed that risks with low uncertainty and impact are the ones familiar from applied science for which traditional mathematical tools of risk analysis are most appropriate. Risks with medium but not high uncertainty and/or impact are in the domain of "professional consultancy," which "uses science; but its problems and hence its solutions and methods, are radically different" (Ravetz 2006, p. 276). The label "postnormal science" applies to situations characterized by high uncertainties and/or high stakes.

Renn and Klinke (2004) similarly use this map with axes denoting uncertainty and impact and identify several areas on that map that delineate qualitatively different risk situations, though these depart from Funtowicz and Ravetz's three areas on the map by being more fine-grained: For example, the points in the map where the probability is low but the potential harm is great, they call "Damocles" risks, named after the Greek king who according to the legend had a sword suspended above him by a thin piece of string (the analogy being that the probability of the string breaking at any one point in time is low, but when it happens, the outcome is rather dramatic, at least for Damocles). Points with high probability and high harm they call "Cassandra" risks, after the Trojan prophet who knew about the fate of the city but whose warnings were ignored. Hovering more in the background is a larger area of the map, where we are not very knowledgeable about the event's probabilities or its outcomes ("Pandora" risks).

Brian Wynne introduced his classification of risks as an improvement on the Funtowicz and Ravetz (Wynne 1992) classification which defines postnormal science. Like Stirling, Wynne sees "risks" as situations where the outcomes and the probabilities are well known and quantifiable. Uncertainties are present when "we know the important system parameters, but not the probability distributions" (p. 114). By contrast, the next level, "ignorance," is more difficult to define: "This is not so much a characteristic of knowledge itself as of the linkages between knowledge and commitments based on it" (p. 114). It is "endemic to scientific knowledge" (p. 115), because science has to simplify what it knows in order to work within its own methods. Finally, "indeterminacy" is seen as largely perpendicular to risks and uncertainties, because it questions the assumption on the causal chains and networks themselves. Thus, indeterminacy can be a huge factor in a particular situation even when the risks and uncertainties are judged to be small.

I am sympathetic to Wynne's classification because it recognizes that both quantifiable types of uncertainties as well as the less tangible deeper uncertainties are present at the same time in some situations and thus not mutually independent, which is a necessary realization away from other schemes such as Funtowicz and Ravetz's map. According to Wynne,

- ▶ Ravetz et al. imply that uncertainty exists on an objective scale from small (risk) to large (ignorance), whereas I would see risk, uncertainty, ignorance and indeterminacy as overlaid one on the other, being expressed depending on the scale of the social commitments ('decision stakes') which are bet on the knowledge being correct (Wynne 1992, p. 116).

However, there are for me still some problems with it. First, and more trivially, is the question of terminology. Like almost every other theorist of risk that comes from the social science side, Wynne and Stirling take "risk" itself to be one of their categories, and then proceed to label the other categories somewhat arbitrarily – this results in a mess of technical definitions that leave no special terminology for the overall thing they intend to classify. We cannot call them classifications of risks (or uncertainties) because risk and uncertainty are already part of the classification system. Moreover, this use of the term risk clashes somewhat with the common definition of risk as a measure that combines uncertainty and outcome. This has not helped that another influential tradition of risk theory embodied by Beck (1992) and Giddens' (1999) work takes risk to mean something altogether more nebulous.

Another concern over Wynne's classification, though, is that the categories seem somewhat hard to pin down, in the sense that indeterminacy, for example, includes the various social contingencies that are not usually captured in conventional risk assessments, but what these

social contingencies are, and how they relate to the other types of uncertainties is not categorically stated. It is not entirely clear, at least to me, where the boundaries lie, or even if there are supposed to be any precise boundaries. Ignorance, he writes, is “conceptually more elusive” and best explained through a lengthy example. All this in effect makes Wynne’s conceptualization hard to explain and therefore possibly ineffective as a tool for bridging the divide between the social and the technical aspects of risks. The inclusion of broad concepts such as social contingencies as well as quantifiability leaves the feeling that Wynne’s categories slice through several useful other distinctions on risk (such as those introduced below, in particular that of Walker et al. 2003). While Wynne’s categories are helpful as a conceptual tool to analyze reactions to risk and identifying shortcomings in conventional scientific approaches to risk that need to be addressed, it remains unclear exactly how they intersect and relate to each other. In a way, our own classification presented below is an attempt to reformulate Wynne’s insights in a way that makes more intuitive sense and which hopefully helps in addressing the question of how Wynne’s categories relate to each other.

Van Asselt and Rotmans (2002) classify risks according to the source of our uncertainty, distinguishing primarily between the two major sources introduced above of epistemic and aleatoric uncertainties (or, in their terminology, uncertainties due to lack of knowledge and uncertainties due to the variability of nature). Uncertainties due to lack of knowledge include, for example, lack of observations/measurements, inexactness or conflicting evidence, while uncertainty due to the variability of nature includes variability in human behavior, value diversity, and the inherent randomness of nature. Aiming to go further than this, Walker et al. (2003) include more dimensions in the classification than merely the source of uncertainty. Thus, they distinguish between location, level, and nature of uncertainty: the location uncertainty can be subdivided between context, model, input and parameter uncertainties, and the final outcome uncertainty. Location uncertainty therefore roughly describes what we are uncertain about, i.e., “where uncertainty manifests itself within the whole model complex” (p. 9). The levels of uncertainty describe the “progression between determinism and total ignorance” and include, in order, statistical uncertainty, scenario uncertainty, recognized uncertainty, and total ignorance. Finally, the nature of uncertainty is, like in van Asselt and Rotman’s classification, mainly about the source of uncertainty, and can roughly be divided into epistemic and ontological uncertainties and subclassified as done by van Asselt and Rotmans.

In this chapter, I hope to be able to add a more inclusive categorization that stays within the spirit of Wynne’s as well as Walker et al.’s (2003) ideas but revolves more centrally around the question of what exactly it is that we are uncertain about, which roughly translates to the “location of uncertainty” dimension in Walker et al. This I will try to use to find interconnections between different literatures on risk. I will argue also that it is useful to apply the scheme to a selection of real-life uncertainties and use it to delineate and make sense of different groups’ varying assessments of a situation because they place different importance on the different objects of uncertainty that are all present to various degrees in all of the cases. I will start by making some preliminary distinctions about risk and uncertainty which will enable us to see where this fits into the various other definitions and classifications of risk. I will borrow Walker et al.’s (2003) idea of different dimensions here, but add that, in our context, these dimensions can best be thought of as different answers to the question on what we want to know about risk.

Firstly, we conceptualize risk as a measure of uncertainty of an event happening times the severity of the outcome. As argued above, this is the usual definition of risk, though it is not

used like this by all commentators, some of whom depart more from Knight's (1971) famous distinction between risks as quantifiable uncertainties versus uncertainties that are not quantifiable, which explains Wynne and Stirling's decisions to put "risk" as one of the categories within their overall schemes. Other writers such as those from the "risk society" tradition (Beck 1992 and Giddens 1999) use risk in a much more vague way which is not so much interested in quantifiable or nonquantifiable or even in the separation of uncertainty and severity of the outcome, but sees it more as the vague possibility that things can go wrong. This is again due to the fact that risk sociologists are interested in different aspects of risk (for example, how increasing awareness and preoccupation of risk affects late modern society). There is therefore not much point in criticizing some work for using vague definitions of risk because, from their point of view, there is simply not that much value added to having a precise working definition of what risk is. However, I hope to be able to show how our distinctions can contribute nevertheless to a better understanding of how the conception of risk that is seen as interesting to sociological and cultural approaches can be compared to other conceptions of risk.

Starting from the definition of risk being a measure of uncertainty and severity of outcome, it is secondly to be noted that neither uncertainty nor severity of outcome are in most cases easily measurable or even definable. Our scheme will leave the very interesting problem of severity of outcome for others to work out and concentrate specifically on the uncertainty aspect of risk.

Starting from the question of "what do we want to know about risk?" we can produce a table of different classifications of risk which are designed to answer that question in different ways. We may, for example, be interested in why we are uncertain, we may be interested in who is uncertain, how it affects individuals or society at large, how is risk represented and how should it be represented, and what is it exactly that we are uncertain about? These are the categories I use below, though there will possibly be more dimensions than those, and other authors may want to divide them differently (Walker et al. (2003), for example, distinguish between levels of uncertainty (whether we take a deterministic position or not) and nature of uncertainty (i.e., uncertainty seen as either aleatoric or epistemological), which I would both see as different sources of uncertainty (we can be uncertain *because* we take an epistemological stance and *because* we have an idealized deterministic situation)).

*Why are we uncertain?* Here we can list classifications that have been made regarding the sources of uncertainty, such as in the scheme of van Asselt and Rotmans, which also relies on the philosophical distinction between epistemological and aleatoric (or ontological) uncertainty described above: we can be uncertain either because of our lack of knowledge or because there is an inherent variability in nature. These two fundamental positions are often seen within the more sociological literature on risk as aspects that different situations of uncertainty can take on, so that, depending on the context, an uncertainty can be either epistemological or aleatoric: "it often remains a matter of convenience and judgment linked up to features of the problem under study as well as to the current state of knowledge or ignorance" (Walker et al. 2003 p. 13). In the philosophical literature, by contrast, it is more often assumed that the distinction is a result of different worldviews: we can, for example, be determinists in our general philosophical outlook, in which case, strictly speaking, all uncertainties are epistemic. In most everyday examples, the boundaries of whether an uncertainty should be considered epistemic or aleatoric seems to be a result of the setup, but the precise boundaries or even existence of the boundary to a large extent also depends on our philosophical stances and background assumptions and knowledge. We can, for example, see the probability of winning

the lottery jackpot with a given set of numbers as purely aleatoric, because even with the most sophisticated current scientific methods, we are some way away from predicting the numbers drawn even if, philosophically, we are strictly speaking determinists who believe that an all-knowing demon could calculate the final result from the initial state. In this example, the existence of probability that is for practical purposes aleatoric even for strict determinists is fairly obvious, though this is not necessarily the case in others. As I will argue below, there are other, epistemic, considerations to be made when we assess the likelihood of winning the lottery.

*Who is uncertain?* The question of the subject of the uncertainty is interesting from the point of view of psychologists or sociologists, who want to know what effect uncertainty has on people or on society at large. Different people respond to uncertainty differently, as shown, for example, in the well-known “white male effect” and similar phenomena discovered by risk psychology research (Slovic 2000). The subject of the uncertainty is also important for policy making since we would need to know how different groups and individuals respond to risks and representations of risk. For example, my current project investigates local opinions on energy infrastructure: to understand the dynamics of risk opinions within the area the infrastructure is being planned, we need to have a more detailed understanding of who the local actors and groups of actors are and how they interact with respect to interpretations of risk. Whether “the public” consents to the infrastructure being built in their back-yard ultimately depends on a complex interplay between local and national politicians, civil servants, the project developers, media representations, local and national NGOs and residents’ interest groups, as well as the individual resident’s understanding which is strongly influenced by, and in turn influences, the other stakeholders. Putting “the public” in scare quotes above is meant to signal that there is no monolithic public, with similar agendas, identities, or worldviews. Understanding who the relevant actors are and how they arrive at their conceptions of risks and how they influence and are influenced by other groups of actors is vital for the analysis of what role risk plays in planning decisions.

*How is uncertainty represented?* Representations of uncertainty can take on different forms, which is again related to where the risk stands and is perceived along the other dimensions. We can, for example, simply deny that there is any uncertainty or risk at all or just concede that there is some, but more or less, undefined uncertainty. If we want (and know more about the situation), we can give a list of possible outcomes, either on their own or with some indication, qualitative or quantitative, on how likely each outcome would be. Should we have chosen a model we think is appropriate, we can give the result of the risk assessment as say a probability, with or without error bars or other representations of uncertainty on that final number.

How we represent risks depends very much on our knowledge of the situation, denying risk is a valid action when we do not know of any, and a simple list of possible outcomes is useful when we lack knowledge of how likely each outcome would be. However, which representation people chose in practice often depends also on what message they want to get across, or even reflects philosophical stances or implicit assumptions made. For example, if we want to make the risk of taking a particular medication look high, we can choose to represent it in relative rather than absolute terms. Similarly, we can give a positive or negative frame: for example, there is technically no difference between saying that “your chance of experiencing a heart attack or stroke in 10 years without statins is 10%, which is reduced to 8% with statins” and “your chance of avoiding a heart attack or stroke in 10 years without statins is 90%, which is increased to 92% with statins” – yet these two formulations have different connotations for

the reader (example taken from Spiegelhalter and Pearson 2008). We can express probabilities in percentages or “natural frequencies,” where research has shown that people are intuitively better able to understand natural frequencies (Gigerenzer 2002). We can produce bar charts, pie charts, “smiley charts” on top of the verbal expressions, and these again convey different impressions of how risky something is. Finally, we can express uncertainties according to our philosophical understanding – if we say that I have a 10% chance of having a heart attack within the next 10 years, that can either mean “10% of people with test results like me will have a heart attack,” or “10% of alternative future worlds will include me having a heart attack.” Again these scenarios while both expressions of the same amount of uncertainty will qualitatively feel different to people, with the second usually seen as the more persuasive way to get people taking their medicines, because it is more personalized (see also Edwards et al. 2001 on the effects of framing risk to patients)

*Responses to uncertainty:* How do people react to uncertainty? Do we, or should we, respond rationally to risk, for example, by doing a cost-benefit analysis to evaluate risks (Sunstein 2005)? Slovic et al. (2004) argue that while analytical and affective are two distinctive ways of reacting to risk, they interplay to produce rational behavior. But maybe even this distinction between affective and analytical needs to be challenged (Roeser 2009, 2010).

On a larger societal level, the risk society literature concerns itself, among of course other things, with how a society responds to risks (specifically our own, late modern – i.e., contemporary Western – society). Here, the issue is not so much about the nature of the risk as such (though it plays a role as I will outline below), or even whether the risks are real or not, but with the role that an increasing awareness of risk plays within late modern society. In particular, they describe the intuitive pessimistic induction through which people have come to realize (or at least believe) that there are always unexpected uncertainties and the possibility of things going horribly wrong with any possible new technological invention (the “unintended consequences of modernity”). Thus, as society has become more reflexive about its own technological achievements, the awareness of risk has become a more powerful driver of social forces than it was previously when risks were more perceived as due to intangible forces of nature rather than consequences of our own society, and therefore modern Western society’s response to risk has become qualitatively different to what it was before.

*Understanding uncertainty:* How people understand uncertainty is a related but somewhat orthogonal issue to the above – this may relate, for example, to the literature of social representations of risk (Joffe 1999; Washer 2004), which uses the social psychological literature of social representations (Moscovici 2000) to characterize how risk issues are perceived and made sense of through associated reasoning – new abstract and intangible concepts as are usually found in topics surrounding risk are conceptually anchored to concepts that are already understood, and thus new concepts are better assimilated into a group’s already held worldviews. Washer, for example, describes through the analysis of newspaper reports of recent new infectious disease outbreaks like SARS or avian flu and how these unfamiliar diseases (and the risks they represent) are being commonly anchored to already understood and familiar diseases (erg. the Spanish influenza outbreak of 1918), or to other aspects, such as vaguely xenophobic expectations of lax health and hygiene practices of the countries of origin; these mechanisms thus place the new disease into different categories of risk than they might otherwise have been perceived if anchored differently.

Hogg (2007) similarly uses a social psychological perspective, social identity theory (Tajfel 1981) to describe issues of intergroup and in group trust, arguing that our social

identities about which groups we belong to effect how we trust the risk statements of others – in-group members are trusted more than out-group members, and even within groups, individuals who are more prototypical in that their characteristics conform well to group norms and values, are trusted more than more marginal members.

Here, we can also list other approaches that are interested in the social construction of risk. Cultural theories of risk such as the influential approach of Mary Douglas (1992) and the risk society argument are relevant here as well, because it is concerned with how societies construct (and thus understand) risks.

*What exactly are we uncertain about?* I left this category until last because this will be my focus in the following section. The object of our uncertainty has been the concern of several classification systems described above when, for example, Wynne talks about uncertainty over causal chains or networks. Similarly, Walker et al. (2003) talk about “locations” of uncertainty, defining that as “where uncertainty manifests itself within the whole model complex” (p. 9), and distinguishing between uncertainty about the context (uncertainties outside of the model), models, inputs, parameters, and the final model outcome. In the following section, I will propose our slightly similar scheme which aims more at a rather general classification of the main types of objects we can be uncertain about which translates somewhat into Walker et al.’s locations of uncertainty, but is aimed at dispensing a too fine-grained classification in favor of one that we feel makes intuitive sense and can help explain different groups’ reactions toward the same risk scenarios.

In slicing the risk literature into these different categories of what they find interesting about risk, I recognize that a lot of work on risk looks at interactions between these different categories: for example, we can be interested in how different representations of risk and different aspects of risk can affect different people or groups of people. But I hope that this way of presenting the risk literature helps make sense of these interactions, and can therefore provide an interesting look into how different aspects of research on risk interlock. Our specific distinction between different objects of risk is itself designed in part to explain different outlooks on risk. In the following sections, I will present the different objects of uncertainty and, following that, explain through a few examples of how objects of uncertainty interconnect with some of the other dimensions of risk in a way which will hopefully give us a fuller description of the different risk scenarios.

## Objects of Uncertainty

---

Our classification (Spiegelhalter and Riesch [in press](#)), somewhat unwise in retrospect, divides the objects of uncertainty into different “levels.” I am calling our decision unwise because this suggests a particular linear hierarchy which may be misleading, but also because other commentators have attached the label “level of uncertainty” to some of the other dimensions of uncertainty outlined above. Specifically, Walker et al. use the term “levels of uncertainty” to describe the spectrum from determinism to “total ignorance.”

We distinguish between three types of uncertainty within the modeling process, and two without. Our use of the term model here is meant to be rather generic. Philosophical and social studies of scientists have shown that the term “model” can be used in varying ways in science (Bailer-Jones [2003](#)), and, thus, generally it does not have the precise definition that it would have in mathematics or statistics. For example, the everyday constructions through which we as

laypeople make sense of risk situations is taken here to be a kind of model as well, since we take our own incomplete information of the world and how we understand things to work and thus gain an understanding of what might happen. The difference between the nonexpert modeling we do in our everyday life and the expert risk assessments is at the end merely a matter of background technical knowledge and competence and levels of commitment, rather than a huge qualitative difference. There is of course much more to be said about lay understanding and construction of risk perspectives, but, for my purposes, it should be enough to use the term “modeling” in an inclusive way that encompasses both expert and lay processing of risk.

By using the term model in this broad sense, I can apply it to different and varying real-life uncertainties and can include the formal mathematical application of the term as well as the more vague, everyday usage of model, in order to achieve applicability of our scheme across a wide variety of real-life risk situations, where precise mathematical or statistical modeling is impossible, impractical, or simply overlooked.

The categories I will present here are not meant to be mutually inclusive, and they will overlap. On the contrary, as I will argue with a couple of examples, in most risk situations, various levels of uncertainty are present at the same time, and our differences of opinion about risks may be due to us giving different importance to different levels.

*Level 1:* Uncertainty about the outcome. The model is known, the parameters are known, and it predicts a certain outcome with a probability  $p$ . An example here is the throw of a pair of dice: Our model is in this case the fundamental laws of classical probability, the parameters are the assumption that the dice are fair and unloaded, and the predicted outcome of, say, two sixes is  $(1/6) \cdot (1/6) = 1/36$ .

This is comparable to the “final model outcome” in Walker et al. On its own, this level of uncertainty exists only in rather idealized situations, as in arguably the example above of the dice. However, this is the level at which we as members of the public are most likely to encounter risk, for example, when we read in a newspaper that “the chances of developing bowel cancer is heightened by 20% if we eat a bacon sandwich every day” (which is a real example taken from the case study further elaborated below). Such clear numbers, in the vast majority of cases, hide the fact that there are additional uncertainties related to the process in which experts arrived at it.

*Level 2:* Uncertainty about the parameters: The model is known, but its parameters are not known (Once the parameters are fixed, then the model predicts an outcome with probability  $p$ ).

This may simply be a lack of empirical information: If only we knew more, we could fix the parameters.

Our concept of uncertainty about the parameters itself hides a variety of different ways in which we can be uncertain about them: We can have fairly good, quantified probabilities about what the parameters should be as they might simply be a matter of getting better information about the system that is being modeled, but, more problematically, we could also be uncertain about how better measurements themselves are achieved, and/or our uncertainty about the parameters can itself only be expressible as a probability distribution, or even only a qualitative list of possibilities, or lastly we might simply have no idea of what the possibilities could be in the first place. Thus, here some of the different dimensions as outlined above intersect with the object of uncertainty: our uncertainty about the parameters can be due to epistemic or aleatoric sources, and it can be represented in different ways.

Unlike Walker et al. we make no distinction between parameter and input uncertainty here, firstly for reasons of simplicity, but also because this more fine-grained distinction is not all that useful when we try to apply our scheme to real-life examples. Similarly, we would also class uncertainties over boundary conditions and initial values into this category as well, all of which may strain the term parameter uncertainty into categories not strictly speaking considered parameters as such – at the end, however, we decided to balance usefulness and simplicity with detail.

*Level 3:* Uncertainty about the model: There are several models to choose from, and we have an idea of how likely each competing model is to reflect reality. Models are usually simplifications about how the world works, and there are often several ways of modeling any given situation.

This is analogous with Walker's model uncertainty, and again, this uncertainty itself can be presented in different ways and may be due to different sources. The way we should represent the uncertainties over model choice is more of a contentious issue and, of course, depends on the precise source of that uncertainty itself. In Spiegelhalter and Riesch ([in press](#)), we advocate a Bayesian approach to compare competing models (after Hoeting et al. [1999](#)). This though will not be everyone's favored approach, which means that, in most situations, we will encounter varying approaches to representations of model uncertainty.

Here, there can be, and frequently are, disagreements between the experts themselves, which means that to the nonexpert public or other consumers of a risk assessment, the uncertainty over the model choice is often related to other factors, such as how much trust they place in the experts to evaluate their model choices honestly or competently, and involves furthermore making a judgment between different experts' assessment when faced with disagreement – however competence and honesty are assumptions that are made only implicitly (only rarely will experts be honest enough to consider their own competence as part of the overall risk assessment – building in an estimation of your own honesty into a risk assessment poses even more problems) and not strictly speaking part of the modeling process. There is therefore is a qualitatively different uncertainty for consumers and for producers of risk assessment, which will be the next level.

*Level 4:* Uncertainty about acknowledged inadequacies and our implicitly made assumptions. Every model is only a model of the real world and never completely represents the real world as such. There are therefore inevitable limitations to even the best models. These limitations could arise because some aspects that we know of have been omitted, or because of extrapolations from data or limitations in the computations, or a host of other possible reasons. Similar in a way to Wynne's concept of "indeterminacy," this is about questioning the assumptions we make, for example, about the validity of the science itself, and thus goes slightly perpendicular to the problems of choosing the models and parameters. These include the "imaginable surprises" (Schneider et al. [1998](#)), that is, things we suspect could occur but about which we do not know enough to be able to include them in the model.

As outlined above, this is where the question of trust comes into force as these are factors that are implicitly not assumed to matter in the risk assessments but not (or rarely) part of it. Similarly, there are always many assumptions about the world that have to be made and that are not part of the modeling process because they are assumed for one reason or another. For example, the risk referred to above of eating too many bacon sandwiches relies not only on the empirical and theoretical studies performed in the analysis, but also on the accumulated medical knowledge about cancer that was taken as given within the risk analysis. Any error

within the fundamental scientific background assumed in a model that is supposed to reflect the real world albeit simplified will make that model less reliable. Therefore, uncertainties in our assumptions and scientific background knowledge are also inadequacies that are acknowledged but not usually part of the modeling process itself. At the same time, these are inadequacies in the process that are at least acknowledged in some way even if not particularly acted upon.

Dealing with acknowledged inadequacies can be done through informal, qualitatively formulated acknowledgment or listing the factors that have been left out of the model, or of course simple denial that there are any in the first place.

*Level 5: Uncertainty about unknown inadequacies:* We do not even know what we don't know. This particular type of uncertainty was made notorious through Donald Rumsfeld's famous speech on "unknown unknowns":

- ▶ There are known knowns. These are things we know that we know. There are known unknowns. That is to say, there are things that we now know we don't know. But there are also unknown unknowns. These are things we do not know we don't know (Rumsfeld 2002).

There are as yet not very many formal approaches to unknown unknowns in the risk literature, though the concept has been well known for a while – for example, Keynes wrote that about some uncertainties, "there is no scientific basis on which to form any calculable probability whatever. We simply do not know" (Keynes 1937). Long before Keynes and Rumsfeld, however, a concept similar to unknown unknowns was introduced by Plato through the famous "Meno's paradox": How can we get to know about something when we are ignorant of what it is in the first place? (Sorensen 2009). It is also related to Taleb's concept of "black swan events" in economics (Taleb 2007) which are events that were not even considered but which, due to their high impact, have a tendency to completely change the playing field, and which is one of the concepts he used to warn about (what turned out to be) the 2008 world financial crisis.

These inadequacies are difficult to deal with formally or informally because we don't really know what they may be, and we are constrained in a way by the limits of our imagination of what could possibly go wrong – Jasanoff (2003), for example, identifies lack of imagination as one of the factors limiting our knowledge for proper risk assessments in postnormal science (p. 234).

Responding to unknown unknowns is naturally very difficult because by definition we do not know what they are. We can however acknowledge them through simple humility that it is always possible that we are mistaken, as demonstrated by Cromwell's quote in the epigraph. Another way is to brainstorm every possibility we can think of and letting our imaginations go wild. This approach is of course never going to be able to cover everything that could go wrong and will therefore not eliminate unknown unknowns.

A slightly more formal way of responding to unforeseen events is the introduction of "fudge factors," for example, in bridge or airplane design, where we design the structure to be a bit stronger than even the worst case scenarios that we could think of require – though even then there is always the conceivable possibility that something worse may happen.

These levels in a way relate to different concerns of different disciplines – who are after all interested in different aspects of risk. For example, the traditional mathematical and philosophical problems of probability theory are mostly concerned with level 1 uncertainty. Statisticians are mostly concerned about level 2 and 3 uncertainty, that is, finding the right model

and, within that model, adjusting the parameters appropriately. It seems unfortunately that uncertainties on which we cannot have a particular mathematical handle on are so often ignored by statisticians and risk modelers – often probably for the pragmatic reason that there simply is not much they can say about the higher levels with the mathematical tools of their trade. Shackley and Wynne (1996), for example, write that in their study of policy discourse on climate change, policy makers were concerned about the validity of the models, while the scientists themselves never even considered that to be an issue, but were instead more concerned about measurement errors within their models. This is to an extent an unfair generalization. An informal survey of technical abstracts from a recent Carbon Capture and Storage conference (Riesch and Reiner submitted) has shown that while model uncertainty is not generally discussed, it does occasionally get mentioned, alongside even an occasional awareness that there are uncertainties associated with unmodeled or unmodelable inadequacies. Nevertheless, worries about model inadequacies were certainly not a prevalent concern among the scientists and risk modelers.

This expert discourse unfortunately distorts the way we perceive particular risks because higher level uncertainties still exist. This may lead to situations like the ones described by Taleb (2007) when he writes about economists having forgotten that unforeseen out-of-the-blue events can occasionally happen and completely mess up our predictions – the sort of events he calls “black swans” if they also have a high potential impact.

The risk society approach of Beck and Giddens is talking mostly about levels 4 and 5, where it is hypothesized that late modern society is living with the increased realization that unmodeled and unmodelable risks are pervasive, and that even if we had some kind of handle on them, there is always the possibility of completely unforeseen events, what Beck calls the “unintended consequences” that he mostly associates with new technology, but which need not necessarily be tied in with it. In Beck’s characterization of late modern society, we have now become accustomed to the realization that despite the best risk modeling of science and engineering experts, technological innovations and advances always have unforeseen consequences, completely left-field occurrences that the original evaluations failed to take into account – in other words, we now know that we live with level 4 and 5 uncertainties all around us. In a way, it matters less to the sociological literatures whether these risks are real or not, but the mere realization that they do happen affects the way late modern society evaluates technological progress and ultimately, itself. Beck’s work has been criticized for ostensibly being about risk, but not quite understanding the concepts of risk analysis and probability (Campbell and Currie 2006), though this slightly misses the point because, within this scheme, it is not really the nature of risk that is important, but responses to it.

As I have tried to argue above, the different disciplinary approaches to risk intersect in different ways – not only do they find different objects of uncertainty important, but they are also interested in different topics among the other dimensions. However, I have not yet found a comprehensive way of translating between the different approaches, and my categorization between different dimensions of risk is meant to solve this. In particular, I feel that the objects of uncertainty dimension which I presented here in more detail can be an important perspective with which to analyze different risk situations in a way that makes sense to the different disciplinary approaches. In the following section, I will go through several examples to illustrate what this perspective can show how all levels of uncertainty are present in most situations involving risks or uncertainties. In particular, I am interested (as a sociologist) to explain how different groups’ perceptions of essentially the same scenario can differ

so dramatically: because through their background experience, assumptions, and worldviews, they will attach different importance to the different levels described above. One important departing point therefore is my assertion that all the levels are present in every risk situation, and that the relative importance that is attached to them depends on who is mulling over it, and I will argue for that below. This seems to be more important on the objects of uncertainty dimensions more than on some of the others, and therefore I feel concentrating on these will help us bring about a more comprehensive way of translating between the various risk literatures, as these will be interested in different objects of uncertainty within each situation.

## Examples

---

In this section, I will explore how these five levels of uncertainty can help explain what happens in various real-life cases in which risk, perceived or real, is a factor, and how our concept of the levels explain different perceptions and how this can lead to the communication difficulties between groups with different perspectives (say between proponents and opponents of carbon capture in the third example). In each of these cases, all five levels of uncertainty are present, though they are differently important and relevant depending on the example.

### The Lottery

I will start with a situation which traditionally is seen as less problematic because it seems to rely only on outcome uncertainty. In a typical national lottery, such as in the UK, there are 49 balls, and each week, 6 balls are drawn; people who have chosen all six correctly win the jackpot. While the exact rules of how much you win are more complicated (depending on the lottery), the case is at least on the surface clearly of level 1: The model is known, the parameters are known, there are no known inadequacies in the model; all the uncertainty that remains is the probability predicted by the model.

This however does not mean that there are no uncertainties present of the other levels, they are simply more hidden and seem less relevant. Level 2 uncertainty concerns the uncertainty of the parameters. In this case, one of the parameters that we have assumed were fixed concerned the individual probability for each ball to come up, thus the question essentially revolves around whether the lottery machine is fair. This is of course a question that we *should* be asking ourselves when we play the lottery, though we rarely do because we trust the authorities that set up the game. As soon as that trust is lost however, level 2 uncertainty comes to the foreground in our evaluations of how likely a jackpot win is. But this is also an empirical question – for the regulator to make sure that the parameters are what we assume them to be, the equipment is regularly checked, and therefore even if we trust the operators to run a fair game, there is still residual empirical uncertainty over the measurements performed during the equipment checks.

Level 3 and 4 uncertainties, in this case, are less likely to bother us because the situation is relatively simple. We, thus, do not really have competing models with which to describe the lottery: unlike in the examples below, where we have situations for which we need a model to describe it, in this case, we start with the model and set up the reality to fit it – that is, after all how the game was constructed. Therefore, in this case, we have a lot of confidence that

the mathematical model we use to describe the game is accurate and not likely to be replaced by one that reflects the situation better.

This however does not necessarily reflect the situation from the point of view of the *consumer* of the lottery – given that the rules of the game are published but the precise probabilities for a given type of win are not necessarily, we have to make our own calculations, and, for the mathematically less able among us (such as myself), there remains the very real possibility that I have made a mistake in estimating my chances of winning. Again the lottery is a pretty simple situation where even I will not have many difficulties; however, the same cannot be said about other games of chance such as blackjack where there is no real model uncertainty from the point of view of an able calculator, but where for the average player, the probabilities are very much subject to uncertainties over mathematical ability. The trust that we have in the operator mentioned above to demonstrate why our parameter assumptions may be wrong is itself a frequently *unacknowledged* inadequacy: The probability that the operator is cheating is, even if somehow quantifiable, rarely part of the model (which in turn makes out the parameter uncertainty to be only dependent on empirical questions) used to estimate the probabilities of winning or the expected pay-out. Yet again, for the consumer who may have a different estimation of the trustworthiness of the operator, level 4 introduces an unmodeled uncertainty, and their estimations of this uncertainty will be different according to background knowledge and assumptions.

Considering completely unexpected scenarios now, maybe the machine could blow up during the draw, invoking maybe the need to refund punters – again this would affect the probability of winning overall in a slight way. Or the operator could be declared bankrupt, in which case, it is not clear whether there would be refunds at all, and the issue would probably only be solved on a case by case basis depending on the whims or political pressure of put upon the government as is the case when other companies fold (even though customers will usually not get refunds if a company goes bankrupt, there are often cases, such as tour operators, where political intervention may make an exception). The possibilities here are of course only restrained by my imagination and as soon as I formulate them they are not strictly speaking unknown unknowns. However, the relative ease with which we can conjure up scenarios which are not foreseen at all points toward a large background level 5 uncertainty which cannot be eliminated completely or even adequately estimated through better modeling.

These considerations I hope demonstrate that even in seemingly very clear situations that are not usually assumed to be subject to other than level 1 uncertainties, our estimation of the uncertainties rely to a large extent on our trust in the operator, our background assumptions and mathematical abilities, and these differ from person to person.

## Saving Our Bacon

What exactly does it mean when we are informed that we are facing an increased risk (by 20%) of bowel cancer if we eat more than 500 g of processed meat a day (WCRF 2007a)? Again, I will hope to demonstrate here that in this claim, there are several levels of uncertainty interwoven because, depending on which perspective we take, we can evaluate the uncertainties of different objects in varying ways. Therefore, making sense of that claim will involve untangling them. (Incidentally, the lifetime risk of bowel cancer is estimated in the report as 5%, which raises to 6% when we eat a lot of red meat a day. In relative terms, the increase of risk is 20%, while in

absolute terms it is 1%. The fact that the WCRF chose to present the more scary relative increase in their press strategy, rather than the more informative absolute increase, tells us a lot about their communication priorities, see also Riesch and Spiegelhalter 2011).

The claim above is based on a meta-analysis performed by the World Cancer Research Foundation of various published trials that investigated the incidence of bowel cancer among people who consume a lot of red meat versus those that do not (WCRF 2007a). One level of uncertainty therefore involves what the studies, as aggregated by the accepted rules of how researchers should do meta-analyses, tell us about eating processed meat: the model is known (in this case, the rules involved of doing the analysis, as well as the rules of the individual studies aggregated in the meta-analysis), the parameters are fixed (in this case the empirical evidence), and together they predict the outcome, bowel cancer, as 20% higher than without the consumption of processed meat. This is the level of uncertainty at which the WCRF communication strategy operated: Our science has found that the risk is p, and that is what the public should know about red meat (as suggested by the WCRF press strategy; WCRF 2007b; WCRF 2007c).

However, especially when looked at from the perspective of the reader of the report, the other levels of uncertainty are there in the background as well and have been emphasized by some of the other actors in the debate: Level 2 uncertainty is, to a certain extent found in the report itself, as this represents the empirical uncertainties surrounding each individual study in the meta-analysis (i.e., fixing the parameters through empirical data): These empirical errors have been aggregated, and since the meta-analysis involved lots of different studies, the overall error has been reduced, and this level of uncertainty is represented through the use of error bars in their charts. Error bars of course did not make it into the verbal communication that accompanies the study's conclusion; instead, the information about uncertainty here is formulated qualitatively: The report distinguishes between the evidence being “convincing,” “probable,” and “limited.” In the final communication of the report, the inherent experimental error, the level 2 uncertainty, was not quantitatively included and could possibly be said to be relatively low. There were though a small qualitative indicators in the wording of the press release:

- ▶ There is *strong evidence* that red and processed meats are causes of bowel cancer, and that there is no amount of processed meat that can be confidently shown not to increase risk (WCRF 2007c, my emphasis).

While level 2 uncertainty has been addressed in the report, if only qualitatively as a way of showing some caution in interpreting the results, level 3 uncertainty posed more problems and was the sort of uncertainty that the expert critics of the report have focused on: This is uncertainty surrounding choosing the model itself. In this case, that translates to the controversy of how the meta-analysis was done, and specifically which studies were included in the analysis. Critics of the report have pointed out that the meta-analysis has left out many individual studies that, if included, would have given the whole analysis a different result. Whether or not there was much merit in these criticisms, they at least demonstrated that no amount of certainty in the analysis itself can remove the uncertainty inherent in choosing the model. The methodology of meta-analysis in general, while an established tool within medical research, is nonetheless not without its critics, and again, though I will not comment on whether these criticisms have much merit, they demonstrate that even within the expert community there are differences of opinion, and therefore, especially for nonexpert bystanders

like me, there is an additional uncertainty over whether the whole methodology used by the report is sound in the first place.

This is compounded by level 4 uncertainty because even if we did have certainty over which studies should have been included, and whether meta-analysis in general is the best way to pool the results from these studies, there is still residual uncertainty about the scientific background assumptions underlying the study, which relies to a large part on previous medical knowledge on for example cancer which is seen as well-established and therefore not considered a factor to be included in the model at all. This is to an extent not too much of an inadequacy because the study being an empirical evaluation of several selected trials and observational studies does not rely much on previous medical knowledge; however, it *does* rely on previously established medical and scientific knowledge and assumptions that these methodologies are a valid way of establishing knowledge. Again though it is not my place to comment on whether there is any merit to these criticisms, that is a criticism that certainly has been made, not so much by medical experts, but by alternative health practitioners who reject a large amount of otherwise established medical knowledge and methods. For the nonexpert bystander, again, the situation is that of competing groups who both claim to have expert status and who have different opinions about what the study shows and can even in principal be expected to show. This is then a different level of uncertainty altogether for the consumer of the report.

Added to that, there are other implicitly made assumptions in the report which relate to the honesty and competence of the researchers themselves. Of course, we can't expect them to take these concerns seriously as an additional uncertainty in their own science, but these are not assumptions that can automatically be assumed by the reader. Both of these two levels of uncertainty (3 and 4) were emphatically not voiced in the official communication by the WCRF, which is understandable because they would have cast doubt on their own experts' judgment. However, they were certainly voiced by the critics: Level 3 uncertainty, as shown above, was the expert critic's response, of fellow medical researchers who accept the general methodology but object to the way it was performed in this instance, while level 4 uncertainty is more usually the response of critics who disagree with meta-analyses generally or who distrust or disbelieve some of the assumptions that medical research takes as established (these are not very influential among medical researchers, but have some influence among alternative medicine campaigners).

Finally, there is level 5 uncertainty: We may be completely wrong footed about the risks of processed meat – maybe the results of all the studies were a systematic error in the design of contemporary medical studies that we do not know about? Maybe something even more exotic has gone wrong? Admittedly, in this particular case, it is quite hard to imagine possible level 5 uncertainties, but this is of course the nature of this level of uncertainty as I have defined it by us not having any handle on it, and not even ever having thought of the possibility. Accordingly, giving a numerical estimation of this level of uncertainty is impossible. Beyond gut instincts, we cannot even tell if it is likely or not likely that something is fundamentally wrong with our conceptions of the problem. In the next examples, I will show that while in this case level 5 uncertainty is not much incorporated in the current thinking about the subject, in many other situations involving risk, level 5 uncertainty can be central.

The complications and disputes involved in this case are connected with the protagonists talking about different levels of uncertainty: The WCRF experts talked about level 1 uncertainties in their take-home message to the general public, while at least acknowledging level 2 uncertainties when talking among themselves and in communication with other experts.

Level 3 uncertainty is the level at which the expert critics attack the report, while the less listened to nonmedical critics attacked it at level 4. Meanwhile, level 5 uncertainty looms menacingly in the background. Some of the media interpretation and discourse about the WCRF study and its press releases can be found in Riesch and Spiegelhalter (2011).

## **Carbon Capture and Storage**

Carbon Capture and Storage (also referred to as Carbon Capture and Sequestration), or CCS, is a technology designed to reduce carbon emissions from fossil fuel burning power plants by capturing the CO<sub>2</sub> through various processes and storing it underground in depleted natural gas reservoirs or other suitable storage sites. The technology is seen by its proponents as an important and technically feasible way to lower carbon emissions because it relies on already fairly well-known mechanisms. While it is admittedly only planned as a relatively short-term solution to be deployed while renewable energy sources are being developed further, it solves some of the problems of “technology lock-in” that could happen if we concentrated only on a few favored energy sources such as solar- and wind-power which we have no guarantee yet that they will be deployable at a large enough scale to reduce carbon emissions in time to avert catastrophic climate change. Therefore, it is seen as part of a necessary portfolio of energy technologies that needs to be included if we want to avoid putting all our eggs in one basket (as argued for example by the influential Stern report, Stern 2007). Further benefits of the technology include more security in the energy supply because it would make burning coal an environmentally sound energy option again and, therefore, reduce the dependency some countries with large coal resources like the UK have on foreign gas imports. A more environmentally appealing further benefit of CCS is that when the technology is developed far enough, it can be used in conjunction with biomass burning power plants and therefore represents one of the few currently technically feasible ways of *removing* carbon from the atmosphere.

Despite these advantages, CCS has many opponents principally among the environmental community, who argue that it merely propagates our dependency on fossil fuels and drives funds away from developing more promising energy technologies which need to be developed anyway because even proponents of CCS see it only as a short term solution (Greenpeace 2008). Finally, one objection to CCS which threatens to be a show-stopper is the safety risks to local people and the local environment that are posed by possible CO<sub>2</sub> leakage from the storage reservoirs and the pipelines that transport the CO<sub>2</sub> from the power plants to these sites. It is these safety risks of CCS that I will concentrate on here; however, the other arguments for and against CCS are relevant here because it is our background worldviews, knowledge, and assumptions which color the way we perceive specific risks. One immediately obvious example of how our background knowledge may color our perception of the risks of the technology concerns our knowledge (and uncertainties within that knowledge) of the toxicity of CO<sub>2</sub>. While CO<sub>2</sub> is not, in fact, neither toxic nor flammable (it does, however, act as an asphyxiant, and therefore still represents a potential though somewhat lessened danger to people living near leakage sites), public opinion surveys on perceptions of CCS have shown that worries over CO<sub>2</sub> are very much in the forefront of public safety concerns (Itaoka et al. 2004; Mander et al. 2010).

Level 1 uncertainty in this case is the final number of the risk assessment, which is usually the basis on which politicians or energy companies would claim that experts find the technology to be very low risk. These numbers are arrived at through models which make of course

several assumptions. A general model of how carbon storage works depends very much on local conditions if we want to arrive at numbers for any particular reservoir and the surrounding area. The local conditions vary in great detail, and therefore experts who perform risk assessments of prospective sites need to investigate them very closely – what are the exact geological formations the CO<sub>2</sub> would be stored in, what are the properties of the cap-rock formations that are needed to keep the CO<sub>2</sub> from traveling up, are there seismic fault lines and if so how would they affect the storage, how many man-made injection wells are there, and how exactly are they going to be sealed once the CO<sub>2</sub> is injected, what is the general three-dimensional shape of the landscape above the possible leakage sites (since CO<sub>2</sub> is heavier than air, there is a chance that it might stay if it leaks into a valley and thus cause greater potential health risks). All these are in a sense parameters that need to be put into the general models if we want to arrive at a final number of expected deaths per year. All of these are subject to their own uncertainties, either because of potential measurement error, or even a more general lack of understanding of the local conditions which in practical terms can only be estimated.

At level 3, there is the choice of general model. In the case of CCS, there is still some argument over whether models developed by the gas and oil industries are really applicable to the storage of CO<sub>2</sub> (Raza 2009). There are also potential debates to be had over precisely what statistical methods should be used and their applicability. Writing in a Dutch popular science magazine article about the proposed (now canceled) CCS storage site under the town of Barendrecht near Rotterdam, Arnoud Jaspers felt that there is some additional uncertainty over model choice when he interviewed modeling experts, for example, some of the models simply did not take into account the three-dimensional structure of Barendrecht and therefore arrived at unreliable scenarios of what would happen should CO<sub>2</sub> leak (Jaspers 2009, 2010). As this shows, not every relevant bit of information makes it into all the models, and there is therefore some uncertainty about which model would be the best to use.

Furthermore, there are other things that are not considered in any model because we simply do not know enough about them or their relevance to be able to model them, these then are the acknowledged inadequacies we term level 4 uncertainties. A recent draft guidance document by the European Union on the implementation of Directive 2009/31/EC concerning geological storage of CO<sub>2</sub> (EU 2010), for example, divides the types of risk expected from reservoir leakage to be “Geological leakage pathways,” “leakage pathways associated with man-made systems and features (i.e., wells and mining activities),” and “other risks such as the mobilization of other gases and fluids by CO<sub>2</sub>” (p. 31). While the first two are routinely part of the models, the “other” category provides more of a problem because, other than listing some possible scenarios that can only to be considered on a “case by case” basis, there is not that much additional analysis that can be introduced. Therefore (as a brief glance through the technical papers on CO<sub>2</sub> storage at the GHGT10 conference has shown – discussed in more detail in Riesch and Reiner submitted), most risk models of CO<sub>2</sub> storage consider leakage pathways along geological fractures or man-made boreholes but do not as such feature other possible leakage pathways either because they are judged to be not very important or, more worryingly, not enough is known about them to include them in the models.

Lastly, there are level 5 unknown unknowns which are not part of the modeling process, not because we do not know enough about them, but because we do not know about them at all. Giving a concrete example is, again, impossible since simply by thinking about them they become known unknowns. However, there are scenarios that can be imagined by the public

that are never even considered in the expert literature. For example, the cover illustration to Jaspers' (2010) popular science article on CCS features a huge "blow-out" scenario with a vast amount of CO<sub>2</sub> escaping explosively and destroying large parts of Barendrecht's neighborhood with rescue helicopters hovering around the scene like tiny flies in relation to the explosion; painting the picture of carbon storage as a huge shaken soda bottle bubbling menacingly underneath the town which will explode spectacularly as soon as there is any kind of leak. This scenario was emphatically not considered in any risk assessment partly because such an explosion would be contrary to anything we know about the behavior of stored CO<sub>2</sub>, and it is probably fair to say that therefore it is not included in risk assessments nor even considered as a known inadequacy of the models. Nevertheless, this scenario of *something* going horribly wrong somehow is a valid concern especially for those who do not possess the expert background knowledge to adequately judge it as so unlikely as to not even be worth including in the model.

This rough overview on the risk debates on CO<sub>2</sub> storage is meant to show that there are very much different perspectives we can take on the risk of CCS, and which ones we put most stock in depend on our background knowledges and ideologies. As in the red meat and cancer example above, different actors in the debate have emphasized different types of uncertainties in this case. Energy companies like Shell or BP who are developing CCS projects, as well as those politicians who are keen on promoting it take comfort from the fact that final risk assessments put the safety risks of the technology as very low, and in much of the industry communication literature on CCS, it is these figures that get mentioned rather than any more technical discussion surrounding how they were acquired. Literature from environmental groups on the other hand see the uncertainties in a different context by highlighting the potential of measurement errors in local evaluations, casting doubt on the modeling processes involved (for example, Greenpeace 2008).

## Climate Change

This leads us to a final very brief example because the one thing that complicates debates about CCS is its relation to the mitigation of climate change. Man-made climate change presents a particular problem, not because there are by now any doubts left that it is happening, but because the forecasting of how bad it will be under various scenarios is a very imprecise business. Scientists who try to predict possible climate futures almost always start by admitting that they are working from one particular model, and that there are several that we know we could use instead that give different estimates, and we rarely have any good handle of working out how likely each model is to reflect reality. In fact, it is often acknowledged that we know so little and climate and weather patterns are so complex that there are always possible factors that we do not even know that we do not know about which can take us by complete surprise.

Uncertainties in climate change modeling are thus dominated by levels 4 and even 5: We have several models to choose from, but not much knowledge on how well they are doing their jobs, and even then we are well aware that our forecasting is hostage to completely unforeseen things as well. Social science research on climate modelers themselves has shown that there is a wide range of expert opinions on the best modeling process and that, moreover, experts themselves will have an unreliable estimation as to the possible shortcomings of their own models (Lahsen 2005). Even then, if they have an adequate estimation of the reliability of their

models, experts find it hard to communicate them, especially when these estimations cannot be quantified (Hillerbrand 2009). Therefore, yet again, for the nonexpert observer of these debates, the most important uncertainty is not over which model is best, but over which expert to trust most, and because there is so much disagreement and unreliability of the experts' own assessments of their models, this uncertainty weighs more for the nonexpert than for the experts.

## Further Research

---

By conceptualizing uncertainty along the various dimensions introduced above we can gain an appreciation of what the different disciplines find interesting about risk and, hopefully, find how they interconnect. The risk society literature, for example, as I outlined above is interested in different aspects of risk than the risk management literature. Considering the particular dimension that I think is most important in my map, the objects of uncertainty, gives us an idea of how and why different people estimate uncertainties differently even when presented with the same information and furthermore shows how different disciplines can themselves study risk from different perspectives.

This is therefore more useful for sociological research than the otherwise admirable combined system of van Asselt and Rotmans (2002) and Walker et al. (2003), whose scheme was designed primarily for use by experts in integrated assessments. Unlike Wynne's and Stirling's systems, which were designed for sociologists to understand and analyze different reactions to uncertainty, my map and classification tries to be more inclusive and meticulous in teasing apart different aspects of uncertainty, while hopefully still being simple enough to be useful for gaining an immediate and intuitive understanding of why and how opinions on risk so often differ. This can then become a useful tool when social scientists communicate the problems of the social contexts of risk to technical experts – this is after all a role performed very often by social scientists who have been funded by scientific research institutions and funding agencies “to look at the social side of things,” but who often struggle to make the social insights relevant and intuitive to the technical experts they work with. This is a main reason why we (Riesch and Reiner submitted; Upham et al. 2011) use this framework in the study of risk opinions on energy infrastructure (CCS and biofuels, respectively), to so far very positive reactions from the technical communities. This approach therefore tries to marry the socio-logical usefulness of Wynne with the technical relevance of Walker and van Asselt and their colleagues.

Though the scheme presented here is meant to be illustrative rather than prescriptive, it shows some clear lessons for risk communication strategies. Since, as I have argued here, different people are worried about different aspects of the risks and, in particular, attach the uncertainty to different objects, risk communication strategies often fail to convey the information that people actually find important. While there is no silver bullet with which to persuade people who simply do not trust the experts, or who's understanding of technological risks gives a higher importance to unforeseen events, taking these different perspectives into account will ensure that the conversation at least does not disintegrate into different actors failing to understand each other. In designing a communication tool about the risks of CCS, for example, we may want to pay particular attention not just to the risks as calculated by the risk assessment, but also how it was arrived at, what the uncertainties with the parameters are, what

was the choice of models available, and why was this particular one chosen, what possible inadequacies were not modeled and finally what are the plans for action should unforeseen consequences occur.

Future research will hopefully develop some of the other dimensions along the more detailed level as I have tried with the objects of uncertainty dimension (and as van Asselt and Rotmans have already done with the sources of uncertainty). This would then allow us to construct a more detailed table through which we can map, at a glance, where the different academic literatures on risk lie and intersect and which may help researchers in finding connections and future ideas for more integrated interdisciplinary research on risk.

Work is also underway to develop case studies which apply the objects classification to different risk scenarios. I have summarized here the application to CCS (Riesch and Reiner submitted); furthermore, we are applying it to the problem of indirect land use change for the biomass energy industry (Upham et al. 2011). Furthermore, detailed studies on more diverse risk situations will hopefully be able to tease out more of the potential but also limitations of our scheme.

## References

---

- Aven T, Renn O (2009) On risk defined as an event where the outcome is uncertain. *J Risk Res* 12(11):1–11
- Bailer-Jones DM (2003) Scientists' thoughts on scientific models. *Perspect Sci* 10:275–301
- Beck U (1992) Risk society: towards a new modernity. Sage, London
- Campbell S, Currie G (2006) Against beck: in defence of risk analysis. *Philos Soc Sci* 36(2):149–172
- Carlyle T (1871) Oliver cromwell's letters and speeches: with elucidations. Scribner, Welford, New York. [http://www.gasl.org/refbib/Carlyle\\_Cromwell.pdf](http://www.gasl.org/refbib/Carlyle_Cromwell.pdf). Accessed 1 Sep 2010
- Douglas M (1992) Risk and blame. Routledge, London
- Edwards A, Elwy G, Covey J, Matthews E, Pill R (2001) Presenting risk information – a review of the effects of “framing” and other manipulations on patient outcomes. *J Health Commun* 6:61–82
- European Union (2010) Implementation of directive 2009/31/EC on the geological storage of carbon dioxide. <http://ec.europa.eu/clima/policies/lowcarbon/docs/GD1-CO2%20storage%20life%20cycle%20risk%20management-consultation.pdf>. Accessed 31 Dec 2010
- Funtowicz SO, Ravetz JR (1990) Uncertainty and quality in science for public policy. Kluwer, Dordrecht
- Funtowicz SO, Ravetz JR (1993) Science for the post-normal age. *Futures* 25:739–755
- Giddens A (1999) Risk and responsibility. *Mod Law Rev* 62:1–10
- Gigerenzer G (2002) Reckoning with risk. Penguin, London
- Gillies D (2000) Philosophical theories of probability. Routledge, London
- Greenpeace (2008) False hope: why carbon capture and storage won't save the climate. Greenpeace International, Amsterdam
- Hacking I (1975) The emergence of probability. Cambridge University Press, Cambridge
- Hillerbrand R (2009) Epistemic uncertainties in climate predictions. A challenge for practical decision making. *Intergenerational Just Rev* 9(3):94–99
- Hoeting J, Madigan D, Raftery A, Volinsky CT (1999) Bayesian model averaging: a tutorial. *Stat Sci* 14:382–417
- Hogg MA (2007) Social identity and the group context of trust: managing risk and building trust through belonging. In: Gutscher H, Siegrist M, Earle TC (eds) Trust in cooperative risk management. Earthscan, London, pp 51–72
- Itaoka K, Saito A, Akai M (2004) Public acceptance of CO<sub>2</sub> capture and storage technology : a survey of public opinion to explore influential factors. In: Rubin ES, Keith DW, Gilboy CF (eds) Proceedings of 7th international conference on greenhouse gas control technologies, 31 volume 1: Peer-reviewed papers and plenary presentations, IEA Greenhouse Gas Program, Cheltenham
- Jasanoff S (2003) Technologies of humility: citizen participation in governing science. *Minerva* 41: 223–244
- Jaspers A (2009) Slapen met de ramen dicht. *Natuurwetenschap & Techniek (NWT)* 77(4):24–33
- Jaspers A (2010) The view from technological journalism. FENCO workshop “CCS and public engagement”, Amsterdam

- Joffe H (1999) Risk and “the other”. Cambridge University Press, Cambridge
- Keynes JM (1937) The general theory. *Q J Econ* 51:209–233
- Knight FH (1971) Risk, uncertainty and profit (reprint of the 1921 edn). University of Chicago Press, Chicago
- Lahsen M (2005) Seductive simulations? Uncertainty distribution around climate models. *Soc Stud Sci* 35(6):895–922
- Mander S, Polson D, Roberts T, Curtis A (2010) Risk from CO<sub>2</sub> storage in saline aquifers: a comparison of lay and expert perceptions of risk. GHGT10, Amsterdam
- Moscovici S (2000) Social representations. Polity Press, Cambridge
- Norton J, Brown J, Mysiak J (2006) To what extent, and how, might uncertainty be defined? Comments engendered by “Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support”: Walker et al., Integrated Assessment 4: 1, 2003. *Integrated Ass J* 6(1):83–88
- Ravetz JR (2006) Post-normal science and the complexity of transitions towards sustainability. *Ecol Complex* 3:275–284
- Raza Y (2009) Uncertainty analysis of capacity estimates and leakage potential for geologic storage of carbon dioxide in saline aquifers. Masters Thesis, MIT, Cambridge
- Renn O, Klinke A (2004) Systemic risks: a new challenge for risk management. *EMBO Rep* 5(S1):S41–S46
- Riesch H, Reiner D (submitted) Different levels of uncertainty in carbon capture and storage
- Riesch H, Spiegelhalter DJ (2011) “Careless pork costs lives”: risk stories from science to press release to media“. *Health Risk Soc* 13(1):47–64
- Roeser S (2009) The relation between cognition and affect in moral judgments about risk. In: Asveld L, Roeser S (eds) The ethics of technological risk. Earthscan, London, pp 182–201
- Roeser S (2010) Intuitions, emotions an gut reactions in decisions about risks: towards a different interpretation of “neuroethics”. *J Risk Res* 13(2): 175–190
- Rumsfeld D (2002) Defense.gov News Transcript. <http://www.defense.gov/transcripts/transcript.aspx?transcriptid=2636>. Accessed 3 Dec 2010
- Schneider SH, Turner BL, Garriga HM (1998) Imaginable surprise in global change science. *J Risk Res* 1(2): 165–185
- Shackley S, Wynne B (1996) Representing uncertainty in global climate change science and policy: boundary-ordering devices and authority. *Sci Technol Hum Val* 21(3):275–302
- Slovic P (2000) The perception of risk. Earthscan, London
- Slovic P, Finucane M, Peters E, MacGregor D (2004) Risk as analysis and risk as feelings: some thoughts about affect, reason, risk and rationality. *Risk Anal* 24(2): 311–322
- Sorensen R (2009) Epistemic paradoxes. In: Zalta E (ed) The Stanford encyclopedia of philosophy. <http://plato.stanford.edu/archives/spr2009/entries/epistemic-paradoxes/>. Accessed 31 Dec 2010
- Spiegelhalter DJ (2010) Quantifying uncertainty. In: Paper presented at “Handling uncertainty in science”, Royal Society, London, 22–23 Mar 2010
- Spiegelhalter DJ, Pearson M (2008) 2845 ways to spin the risk. <http://understandinguncertainty.org/node/233>. Accessed 31 Dec 2010
- Spiegelhalter DJ, Riesch H (in press) Don’t know, can’t know: embracing scientific uncertainty when analysing risks. *Philos T Roy Soc A*
- Stern N (2007) The economics of climate change. Cambridge University Press, Cambridge
- Stirling A (2007) Risk, precaution and science: towards a more constructive policy debate. *EMBO Rep* 8(4):309–315
- Sunstein CR (2005) Laws of fear. Cambridge University Press, Cambridge
- Tajfel H (1981) Human groups and social categories: studies in social psychology. Cambridge University Press, Cambridge
- Taleb N (2007) The black swan: the impact of the highly improbable. Penguin, London
- Upham P, Riesch H, Tomei, J Thornley P (2011) The sustainability of woody biomass supply for UK bioenergy: a post-normal approach to environmental risk and uncertainty. *Environ Sci Policy* 14(5):510–518
- van Asselt MBA, Rotmans J (2002) Uncertainty in integrated assessment modelling: from positivism to pluralism. *Clim Change* 54:75–105
- Walker WE, Harremoes P, Rotmans J, van der Sluijs JP, van Asselt MBA, Janssen P, Krayer von Krauss MP (2003) Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. *Integrat Ass* 4(1):5–17
- Washer P (2004) Representations of SARS in the British newspapers. *Soc Sci Med* 59:2561–2571
- WCRF (2007a) Food, nutrition, physical activity, and the prevention of cancer: a global perspective. WCRF, Washington DC
- WCRF (2007b) Landmark report: Excess body fat causes cancer. [http://www.wcrf-uk.org/press\\_media/releases/31102007.lasso](http://www.wcrf-uk.org/press_media/releases/31102007.lasso). Accessed 13 Aug 2008
- WCRF (2007c) Media quotes. [http://www.wcrf-uk.org/press\\_media/quotes.lasso](http://www.wcrf-uk.org/press_media/quotes.lasso). Accessed 13 Aug 2008
- Wynne B (1992) Uncertainty and environmental learning: reconceiving science and policy in the preventive paradigm. *Global Environ Chang* 2(2): 111–127

# Part 2

## Specific Risks



# 5 The Economics of Risk: A (Partial) Survey

Louis Eeckhoudt<sup>1,2</sup> · Henri Louberge<sup>3</sup>

<sup>1</sup>IÉSEG School of Management, Lille, France

<sup>2</sup>The Center for Operations Research and Econometrics (CORE),  
Université catholique de Louvain, Louvain, Belgium

<sup>3</sup>Geneva Finance Research Institute (GFRI), University of Geneva  
and Swiss Finance Institute, Geneva, Switzerland

<i>Introduction</i> .....	114
<i>From the Origins to the End of the 1960s: A Broad Overview</i> .....	115
Expected Utility .....	115
Attitudes Toward Risk .....	117
Risk Aversion .....	117
The Allocation of Risks Among Individuals .....	118
<i>Mean-Preserving Changes in Risk</i> .....	120
<i>Preferences Versus Choices</i> .....	121
<i>Higher-Order Risk Attitudes: Prudence and Temperance</i> .....	122
<i>Relative Risk Attitudes</i> .....	125
<i>Further Research: Applications – Insurance as an Example</i> .....	127
<i>Conclusion</i> .....	130

**Abstract:** This survey provides a brief overview of the treatment of risk in economics, starting from early principles in the 1940s and 1950s and extending until the most recent developments. It shows how original ideas about economic behavior under conditions of risk and earlier definitions of risk aversion were progressively refined to include prudent behavior in risky situations and precise concepts of risk measurement. The survey is partial because it focuses on the mainstream model of economic behavior under risk. It ignores, among other topics, issues raised by the contractual relationships between imperfectly informed agents in markets for risk transfers, as well as behavioral traits (such as loss aversion) often observed among market participants. Besides, it does not cover models of risk management that have become popular in financial mathematics. Some applications of the economics of risk in the insurance domain are however briefly reviewed.

## Introduction

---

Between Daniel Bernoulli's seminal contribution ([1738](#)) and the mid 1940s the economic theory of risk made little progress. An entertaining account of this slow development can be found in Borch ([1990](#)) (see also Bernstein ([1996](#))). Since then, however, the field has been continuously and rapidly expanded and it is now very difficult – if not impossible – to summarize the state of the art in a limited amount of space.

In this survey, we describe some very specific aspects of this wide expansion process. They will be limited to developments made in the mainstream model – the expected utility model (EU) – even though alternative models will be briefly mentioned in section [From the Origins to the End of the 1960s: A Broad Overview](#). We will also limit our presentation to the analysis of individual economic behavior under conditions of risk, leaving aside issues raised by the risk management of corporations and financial institutions and by economic equilibrium in the markets for risk transfer when agents have differential information. These issues are more specifically addressed in domains like corporate finance, financial economics, and the theory of contracts – all domains that draw on lessons from the economics of risk. Notice also that, for lack of space, many other aspects of modern risk theory will not be discussed here. For instance, the notions of “risk measure” and “value-at-risk” (VaR), initially developed in the mathematical finance and actuarial literature (see Artzner et al. [1999](#); Crouhy et al. [2001](#); Föllmer and Schied [2002](#)) and that led to so many theoretical, empirical, and practical developments over the past years, will be left aside in this survey. Similarly, the numerous implications of the economics of risk in more applied fields such as health economics, financial economics, public economics, and environmental economics will not be covered here. We will restrict ourselves to a very brief account of some developments in insurance economics.

Our survey is organized as follows. In the first section we provide a broad overview of developments until the end of the 1960s. These include the principle of optimal behavior under conditions of risk – the maximization of expected utility – the definition and measurement of risk aversion and the fundamental results about the optimal allocation of risks among individuals. While we concentrate in the second section on personal attitudes toward risk, we adopt in the third section a more “statistical” approach and present the fundamental concept of a “mean-preserving increase in risk.” Combining results from the second and the third sections, we discuss in the fourth section the impacts of risk on preferences on one hand

and on choices on the other hand. The fifth section reviews more recent literature focusing on risk attitudes beyond risk aversion, namely, prudence and temperance. The results reviewed in the fifth section lead us to cast a new look in the sixth section on the coefficient of relative risk aversion initially presented in the first section. Our survey ends in section [Further Research: Applications – Insurance as an Example](#) with some comments about applications and a brief account of results obtained in the theory of insurance demand, followed by the [Conclusion](#).

## From the Origins to the End of the 1960s: A Broad Overview

---

Although he never used any of the expressions we are so familiar with today (“declining marginal utility,” “concavity,” “risk premium,” etc.), Daniel Bernoulli set up in 1738 the foundations of the modern economic theory of risk. Using either abstract examples (e.g., the St. Petersburg game) or more real-world examples (the Amsterdam merchant who wants to insure goods to be shipped), he convincingly argued that decision makers (DMs) could attach to a lottery a value below its mathematical expectation contrarily to Pascal and Fermat who had argued that the value of a lottery was given by the mathematical expectation of its consequences. His argument relied on the idea that the utility of gains – not the gains themselves – matters and that utility increases with gains at a declining rate. Quite surprisingly, his so fundamental and deep ideas remained unnoticed by economists for a very long time. It is fair to say that these ideas were not formulated in precise modern economic terms until the works of von Neumann and Morgenstern (1947), Friedman and Savage (1948), Arrow (1965, 1971) and Pratt (1964).

## Expected Utility

---

Thanks to the work of von Neumann and Morgenstern (1947), later extended by Savage (1954), we know under which assumptions of rational behavior a risky situation can be evaluated by the expected utility of consequences. More formally, if a lottery  $\tilde{w}$  – i.e., a probability distribution of outcomes – leads to mutually exclusive outcomes ( $w_1, w_2, \dots, w_s, \dots, w_S$ ) with probabilities of occurrence ( $p_1, p_2, \dots, p_s, \dots, p_S$ ), where  $s = 1, 2, \dots, S$  denotes the state of nature with probability of occurrence  $p_s$  leading to outcome  $w_s$ , its value is given by its expected utility (EU):

$$E[u(\tilde{w})] = \sum_{s=1}^S p_s u(w_s)$$

where  $u$  is a cardinal increasing function of  $w$ .

For a continuous density of outcomes  $f(w)$  defined on  $[a, b]$ , one has:

$$E[u(\tilde{w})] = \int_a^b u(w)f(w)dw.$$

The *expected utility theorem*, as expressed by von Neumann and Morgenstern (1947), states that an individual makes decisions under risk as if he or she maximized the expected value of a *cardinal* utility function of outcomes. For each individual, the utility function is unique up to a positive linear transformation. The theorem is based on a few axioms defining consistent behavior in choices among lotteries:

- (A1) *Preordering*: Considering a set of lotteries  $A, B, C, \dots$ , each pair of lotteries is characterized by a preference or indifference relationship ( $\succsim$ ): either  $A \succsim B$ , or  $B \succsim A$ . Besides, the relationship is reflexive ( $A \succsim A$ ) and transitive: if  $A \succsim B$  and  $B \succsim C$ , then  $A \succsim C$ .
- (A2) *Continuity*: If  $A \succsim B \succsim C$ , then  $\exists q \in [0, 1]$  such that  $B \sim \{(q, A); (1 - q, C)\}$ , where the symbol  $\sim$  means “is indifferent to” and where  $B$  is a “compound lottery.”
- (A3) *Boundedness*: For any lottery  $L$ , there exist bounds  $y$  and  $z$  such that  $L \succ y$  and  $z \succ L$ ,  $\forall L$  where the symbol  $\succ$  means “is preferred to.”
- (A4) *Independence or substitution*: Consider three lotteries  $A, B$ , and  $C$ . Then, for any  $C$

$$A \succ B \Rightarrow \{(p, A); (1 - p, C)\} \succ \{(p, B); (1 - p, C)\} \quad \forall p \in (0, 1]$$

$$A \sim B \Rightarrow \{(p, A); (1 - p, C)\} \sim \{(p, B); (1 - p, C)\} \quad \forall p \in [0, 1]$$

Compared with the axioms needed to represent choices under certainty using an *ordinal* utility function, it turns out that only the fourth axiom is new. This axiom is however necessary to yield the EU theorem. For this reason, it has been the target of criticisms, starting with Maurice Allais (1953) arguing that this axiom goes beyond requirements of rational (i.e., consistent) behavior. In a famous experiment, he showed that a significant proportion of individuals violate expected utility theory when faced with choices where one of the lotteries is the certainty of receiving a large amount (Allais' paradox). His observation has been repeatedly confirmed in numerous experiments since then. Axiom A4 is at stake in observed violations of expected utility theory. The Allais paradox has thus led some scholars to develop competing theories of decisions under risk, not relying on the independence axiom and commonly referred to as “nonexpected utility theory.” The most well known is *prospect theory* (Kahneman and Tversky 1979; Tversky and Kahneman 1992). However, these theories have met with criticism too. None of them has received widespread acceptance so far and none has proved as fruitful and flexible as EU theory in the development of models explaining various features of economic life.

In von Neumann and Morgenstern (1947), the probabilities of outcomes are given and the measurable utility underlying choices may be easily derived, for each decision maker, observing his or her choices. The utility function reflects his or her subjectivity, his or her behavior toward risk. Savage (1954) goes one step further: introducing concepts of states of nature, events, and conditional preferences, and using a different set of axioms he shows that rational choices in risky situations are driven by the maximization of subjective expected utility. The choices reflect the existence of a measurable utility function over outcomes *and* of a subjective probability measure over events. For a clear presentation of the axioms used by Savage, see Dumas and Allaz (1996). The two approaches to expected utility mirror the two concepts of probability: in von Neumann and Morgenstern, probabilities are provided by experience (frequency interpretation), whereas in Savage probabilities reflect a subjective degree of belief. See also Fishburn (1989).

## Attitudes Toward Risk

---

In an early contribution, Friedman and Savage (1948) showed that the shape of the von Neumann–Morgenstern utility function defines behavior toward risk. If the outcome brings utility, the function is increasing: marginal utility  $u'(w)$  is positive for all individuals, whatever their attitude toward risk. The latter is reflected in the second derivative.

- An increasing linear function denotes indifference toward risk: the individual behaves as if he or she maximized the expected outcome,  $\max E[u(w)] \equiv \max u[E(w)]$ . In this case, marginal utility is constant and the second derivative equals zero.
- A concave function denotes risk aversion: the individual prefers the certainty of obtaining outcome  $E(w)$  to the uncertain prospect  $\tilde{w}$  with expected value  $E(w)$ :  $E[u(w)] < u[E(w)]$ . This inequality, known as Jensen's inequality, characterizes a concave function: marginal utility is decreasing,  $u''(w) < 0$ .
- Finally, a convex function denotes a risk-loving individual: risk yields more utility than the certainty of the expected outcome,  $E[u(w)] > u[E(w)]$ . In this case, marginal utility is increasing,  $u''(w) > 0$ .

Besides this definition of risk aversion, Friedman–Savage observed that many individuals purchase insurance contracts, for all sorts of risk, and also buy lottery tickets. In the first case, they behave as risk averters: they pay more than the expected value of loss to avoid a loss; in the second case they behave as risk lovers, paying more than the expected value of gain to participate in the lottery. To solve the paradox, Friedman and Savage proposed a utility function with an inflection point at the level of current wealth, a convex segment following a concave segment. This ad hoc solution has not been retained in the subsequent literature. It is commonly accepted today that risk aversion prevails in society, which leads to justify the use of concave von Neumann–Morgenstern utility functions. Marginal utility is decreasing, an assumption familiar to economists for a very long time. Participation to lotteries by fundamentally risk-averse individuals is rationalized by other considerations such as overestimation of probabilities of gains, pleasure from gambling, and desire to change for a “better life” with more enjoyment, i.e., to switch to a different utility function.

## Risk Aversion

---

In two almost simultaneous, yet independent, papers, Arrow (1965) and Pratt (1964) defined the risk premium, denoted  $\pi$ , as an amount of money a risk-averse decision maker (DM) would accept to pay to replace a random prospect  $\tilde{w}$  by its expectation  $E(\tilde{w})$  received with certainty.

Formally,  $\pi$  is solution to:

$$E[u(\tilde{w})] = u(E(\tilde{w}) - \pi) \quad (1)$$

Using a Taylor series approximation, more particularly appropriate for relatively small risks, it can be shown that

$$\pi \approx \frac{\sigma_{\tilde{w}}^2}{2} \left( -\frac{u''(E(\tilde{w}))}{u'(E(\tilde{w}))} \right) \quad (2)$$

where  $\sigma_{\tilde{w}}^2$  is the variance of  $\tilde{w}$ .

The ratio  $\left(-\frac{u''(E(\tilde{w}))}{u'(E(\tilde{w}))}\right)$  is positive for a risk-averse DM. It is called the coefficient (or index) of *absolute risk aversion*. This index, that we denote  $A$ , has played a central role in the economics of risk, as well as in other fields like finance, operations research, and psychology, to name a few. It turned out to be more particularly important through assumptions made about its behavior when the expected outcome changes. The term “absolute” reflects the fact that  $A$  plays a role in optimal decisions under risk when absolute outcomes are considered, e.g., the chance of a gain or of a loss of a certain amount. Defining the outcomes  $w$  as amounts of wealth, Arrow’s original assumption was that  $A$  decreases in wealth: a wealthy individual is less concerned by the risk to gain or lose \$100 than a poor individual. This assumption of *decreasing absolute risk aversion* (DARA) has been repeatedly supported by empirical evidence and experimental studies (see, for instance, Levy 1994). In financial economics, e.g., it implies that risky assets, like common stocks, are normal goods: the demand for these assets increases when wealth increases.

Arrow and Pratt also developed the concept of *relative risk aversion* linked to a *proportional* risk premium. The coefficient of relative risk aversion is defined as  $R = wA$ , where  $w$  represents the DM’s initial wealth. It turns out that  $R$  is useful when the amount at risk is a proportion  $\alpha$  of the initial endowment, with final random outcome  $\tilde{w} = w(1 + \alpha\tilde{\varepsilon})$ . Arrow hypothesized that  $R$  is increasing with wealth, meaning that a risk-averse DM will invest a smaller proportion of his or her wealth in risky ventures as his or her wealth increases. However, this hypothesis of increasing relative risk aversion (IRRA) has not been supported unambiguously by empirical observations and experimental surveys. Indeed, none of the hypotheses of increasing, decreasing, or constant relative risk aversion could be rejected with high confidence until now (see Meyer and Meyer 2005).

## The Allocation of Risks Among Individuals

A very important concept in economics is general equilibrium. Walras (1874) demonstrated mathematically that, whatever the number of products and markets in the economy, a system of free competitive markets governed by the forces of demand and supply tends to a situation where all markets are simultaneously in equilibrium – the effective demand being equal to the effective supply at the clearing price. Pareto (1896) later showed that this general equilibrium situation is characterized by a remarkable welfare property: it is impossible to increase the utility of one individual without simultaneously decreasing the utility of at least one other individual (Pareto optimality).

The Walras model of general equilibrium is a static model: it has one period and no uncertainty. In the wake of debates about risk and uncertainty during the 1950s, Debreu (1953, 1959) and Arrow (1953, 1964) used the state of nature concept to extend Walras’ model to multiple periods and uncertainty about the future. Debreu (1953, 1959) introduced the concept of state-contingent goods. Goods and services are characterized not only by their intrinsic characteristics, but also by the specific date and state of nature for their delivery. For example, with  $m$  goods and services, two periods, 0 and 1, and  $S$  states of nature at period 1, there are  $m + mS$  markets:  $m$  spot markets and  $mS$  state-contingent forward markets. An umbrella delivered on November 1 if it rains is not the same state-contingent good as an umbrella delivered on the same date if the sun shines. Essentially, the analysis remains the same but it applies to a much larger number of markets, with the same general equilibrium and

optimality properties. The crux of Arrow's (1953) contribution was to show that finance allows to economize on markets. In a monetary economy, the introduction of contingent securities (or conditional claims) leads to the same general equilibrium as the Debreu model with contingent goods. At each period, only  $m + S$  markets are necessary, instead of  $m + mS$ :  $m$  markets for the spot exchange of goods against the numeraire and  $S$  markets for the spot exchange of numeraire against conditional claims on the future numeraire. Instead of purchasing an umbrella delivered on November 1 if it rains, the forward-looking individual purchases a financial contract promising to pay a certain amount at that date if it rains. This amount may then be used to purchase a spot umbrella (or a raincoat). The same result is obtained by economizing on markets: finance favors efficiency. Note also that the two models, Arrow and Debreu, work if there exist complete markets for future delivery:  $mS$  markets for contingent goods in the Debreu version,  $S$  markets for contingent claims in the Arrow version.

A characteristic property of Pareto-optimal general equilibrium is that the ratios of marginal utilities are equalized among individuals. For example, considering two individuals  $i$  and  $k$  in the economy, two future states of nature  $s$  and  $\theta$ , and assuming identical subjective probabilities among individuals, the optimal allocation is characterized by:

$$\frac{u_i'(w_{i\theta})}{u_i'(w_{is})} = \frac{u_k'(w_{k\theta})}{u_k'(w_{ks})} \quad (3)$$

This optimal allocation is obtained either by having a complete system of free competitive markets with each individual maximizing his or her utility under a budget constraint, or by letting a benevolent and omniscient dictator allocate resources by maximizing a weighted sum of individual utilities under constraints of available resources.

Expression (3) is known as *Borch's condition* (Borch 1960, 1962). It has important societal implications. By inspection, it may be seen that if individual  $i$  has an optimal allocation with more wealth in state  $\theta$  than in state  $s$ , the ratio on the LHS of (3) is less than one, by diminishing marginal utility. The expression implies that this must also hold for individual  $k$  on the RHS. As  $i$  and  $k$  are arbitrary, this implies further that aggregate wealth is larger in state  $\theta$  than in state  $s$ :  $W_\theta > W_s$ , where  $W_s = \sum_i w_{is}$ . As a result it must be that

$$w_i = f_i(W) \quad (4)$$

all  $i$ , with  $f_i$  an increasing function.

Expression (4) reflects the *mutuality principle*: at the Pareto-optimal equilibrium, whatever the way to obtain it (free-markets economy or mandatory allocations by the State), it must be that idiosyncratic (individual) risks do not play a role. The allocation of each individual is an increasing function of aggregate resources. If aggregate wealth is higher, everybody is better off; if aggregate wealth is lower, everyone suffers. The individual risks have been canceled by diversification, for instance, using social insurance or a system of complete markets for private insurance. Only the macroeconomic (systematic) risk is left.

This risk cannot be eliminated by diversification. It must be shared among individuals, one way or the other. One possibility is a mandatory distribution of resources, using for instance an egalitarian system. But in a market economy, this allocation is performed by the financial market. Every security in this market is actually a portfolio of Arrow–Debreu conditional claims. By exchanging securities on financial markets individuals and institutions perform changes in the allocation of risks related to future wealth. A simple mathematical derivation

starting from the optimality conditions yields that each individual participates in this process according to his or her absolute risk tolerance (the inverse of absolute risk aversion). It turns out that function  $f_i$  in (❸ 4) has positive first derivative given by

$$f'_i = \frac{\partial w_i}{\partial W} = \frac{T_i^a}{\sum_k T_k^a}$$

where  $T^a = 1/A$ .

Thus, participation in aggregate risk is driven by risk tolerance. Individuals with a relatively higher degree of risk tolerance purchase more risky financial assets. Doing this, they take a larger share of aggregate risk. Compared to more risk-averse individuals, they will lose more if aggregate wealth decreases and gain more if the economy performs well and aggregate wealth increases. To sum up, in the economy, insurance markets perform the diversification of individual risks and financial markets allow the sharing of aggregate risks. Notice, finally, that a particular class of von Neumann–Morgenstern utility functions – the HARA class where HARA means “hyperbolic absolute risk aversion” – yields linear absolute risk tolerance coefficients and therefore linear sharing of systematic risks among individuals.

## Mean-Preserving Changes in Risk

In the early seventies Rothschild and Stiglitz (RS) (1970, 1971) published two papers that nicely complemented those of Arrow and Pratt (AP) discussed above under section ❸ Risk Aversion. While AP had looked at preferences, i.e., properties of a DM’s utility function, RS paid essential attention to the statistical properties of the risk faced by the DM and then established a link between these statistical properties and the utility function.

To start, they raised the following question: when can we say that a random variable  $\tilde{Y}$  is riskier than another one  $\tilde{X}$ ? To make the comparison feasible and independent from other considerations, they assumed that the two variables have the same mean, i.e.,  $E(\tilde{X}) = E(\tilde{Y})$ , hence the term “mean-preserving.” While Pratt had given different plausible definitions of risk aversion and then shown their equivalence, RS did the same for the notion of mean-preserving changes in risk. Here are two possible and apparently different definitions of a mean-preserving increase in risk:

- $\tilde{Y}$  is obtained from  $\tilde{X}$  by adding zero-mean independent risks  $\tilde{z}$  to  $\tilde{X}$ ;
- $\tilde{Y}$  is obtained from  $\tilde{X}$  by taking some weight away from the center of the density of  $\tilde{X}$  and transferring it to its tails without changing the mean (*mean-preserving spread*, or MPS).

Each of these two different definitions makes sense and RS showed that both are equivalent. Besides, they established an important link with preferences that is stated as follows:

- If  $\tilde{Y}$  is an MPS of  $\tilde{X}$ , then every risk-averse DM will prefer  $\tilde{X}$  to  $\tilde{Y}$ .

This result is really important. First, it suggests a very general definition of risk aversion that is based on attitudes toward changes in risk instead of being based on the concavity of the utility function, something specific to the EU model. Further, as will be shown in section ❸ Preferences Versus Choices, this result leads to an interesting distinction between preferences on the one hand and optimal choices on the other. Finally, as will be shown in section ❸ Higher-Order Risk Attitudes: Prudence and Temperance, the RS paper led much later to

the introduction of new concepts such as prudence and temperance which characterize attitudes toward risk beyond risk aversion – hence the term “higher-order risk attitudes.” These new concepts are themselves related to statistical notions such as “downside risk increase” (Menezes et al. 1980) or “outer risk increase” (Menezes and Wang 2005) which are described by Ekern (1980) under the more general term of “ $n$ th degree increases in risk” (see section [Further Research: Applications – Insurance as an Example](#)).

To be fair toward a large body of statistical and economic literature, it is worth mentioning that the notion of MPS represents a special case of *second-order stochastic dominance*, in which the comparison between random variables is made without the restriction of equal means. The notion of stochastic dominance was introduced into economics by Hadar and Russell (1969) and Hanoch and Levy (1969). For a detailed survey, see Levy (1992).

## Preferences Versus Choices

---

Among many other important results, Pratt (1964) indicated in the last section of his paper that, in accordance with intuition, as a DM becomes more risk averse, he or she would engage less in risky activities. This result was confirmed later on in the analysis of many specific situations, e.g., in Mossin (1968) for insurance decisions, or in Sandmo (1971) for production decisions.

Because this result was so natural, a widespread belief developed about the answer to an apparently similar question: given the DM’s degree of risk aversion, will he or she take a less risky position when the return on a risky investment vehicle becomes riskier? Because an increase in risk aversion (for a given risk) or an increase in risk (for a given risk aversion) are both welfare reducing for risk-averse DMs, it was thought they would also have a similar impact on optimal choices. But in fact, it was realized in the mid 1980s that a riskier investment return did not necessarily imply less risk taking by a risk-averse DM. In a sense, the MPS concept as developed by Rothschild and Stiglitz (see section [Mean-Preserving Changes in Risk](#) above) proved to be too general to produce the desired intuitive result.

From this observation emerged a literature that tried to narrow down the MPS concept to a less general concept that would yield the meaningful comparative statics result. The reader is referred to a succession of papers – Eeckhoudt and Hansen (1980); Meyer and Ormiston (1985); Black and Bulkley (1989); Dionne et al. (1993) – in which sequentially less restrictive assumptions are imposed upon the MPS to generate the required result about the observed choices of a risk-averse DM. This literature culminated (and came to an end) with a paper by Gollier (1995) who derived the necessary and sufficient condition that must be imposed on the MPS to induce less risk taking by a risk-averse DM. Since this condition is technical, the reader is referred to the original article or to a survey of this question by Eeckhoudt and Gollier (2000).

While the economics literature on these topics has been developed around the notion of an MPS, a very interesting contribution on a related question in the financial management literature is worth mentioning. Fishburn and Porter (1976) looked at the optimal choice for an investor when the risky investment return benefits from an improvement in the sense of first-order stochastic dominance (FSD), i.e., a “good news” since a random variable  $\tilde{X}$  first-order dominates another one  $\tilde{Y}$  if for any value  $m$ ,  $P_r(\tilde{X} \leq m) \leq P_r(\tilde{Y} \leq m)$ .

Of course, such a good news improves the DM’s welfare (as soon as  $u' > 0$ ), but will it imply more risk taking? As in the case of the MPS, Fishburn and Porter showed that the answer is not straightforward. They then gave sufficient conditions either on the DM’s utility function or on

the nature of the FSD improvement to obtain the desired result. Quite interestingly, the desired result on the utility function is that the coefficient of relative risk aversion  $R$  defined in section [\(6\) Risk Aversion](#) should not exceed unity. We return to this condition in section [\(6\) Relative Risk Attitudes](#).

## Higher-Order Risk Attitudes: Prudence and Temperance

While the assumptions made about the behavior of the absolute risk aversion coefficient  $A$  turned out to be appropriate to discuss insurance or portfolio choice problems, they turned out to be insufficient for the analysis of savings choices. In this field, early papers by Leland (1968); Sandmo (1970); Drèze and Modigliani (1972), and later on Kimball's (1990) contribution stressed the role of the sign of the third derivative of the utility function in the unveiling of a precautionary motive for saving. Since Kimball coined the term "prudence" for a positive value of  $u'''$  and since  $u''' > 0$  generates precautionary saving – as we will show below – it was admitted for a long time that prudence and precautionary saving were "the same thing."

To see this link, let us consider two very simple saving problems with a two-period horizon. In the first problem, current and future incomes ( $y_1$  and  $y_2$ ) are known with certainty. The individual's problem is to choose a current consumption level,  $c_1$ , maximizing intertemporal utility, i.e.,

$$\underset{c_1}{\text{Max}} \ u(c_1) + u\{(y_1 - c_1) + y_2\} \quad (5)$$

To simplify the notation, we assume – without loss of content – that utility is time additive and that the interest rate and discount factor on utility are zero. The optimum is obtained for a value of  $c_1$ , denoted  $c_1^*$ , such that:

$$u'(c_1^*) - u'\{(y_1 - c_1^*) + y_2\} = 0 \quad (6)$$

so that  $c_1^*$  satisfies  $c_1^* = \frac{y_1 + y_2}{2}$ .

This well-known result illustrates "consumption smoothing": thanks to his or her choice of  $c_1^*$  the DM can obtain the same consumption level in each period.

Now, let us introduce risk in the savings problem. This can be done by considering that future income becomes random so that  $y_2$  is replaced by  $\tilde{y}_2 = y_2 + \tilde{\varepsilon}$ , where  $\tilde{\varepsilon}$  is a zero-mean risk.

The optimization problem is now:

$$\underset{c_1}{\text{Max}} \ u(c_1) + E\{u\{(y_1 - c_1) + y_2 + \tilde{\varepsilon}\}\} \quad (7)$$

with a first derivative evaluated at  $c_1^*$ :

$$u'(c_1^*) - E\{u'\{(y_1 - c_1^*) + y_2 + \tilde{\varepsilon}\}\} \quad (8)$$

Comparing [\(6\)](#) and [\(8\)](#) one notices that, by Jensen's inequality:

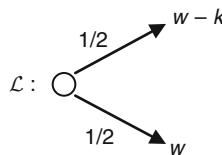
$$E\{u'\{(y_1 - c_1^*) + y_2 + \tilde{\varepsilon}\}\} \geq u'\{(y_1 - c_1^*) + y_2\} \Leftrightarrow u''' \geq 0.$$

If  $u'''$  is positive, it can then be shown that the solution to problem [\(7\)](#), denoted  $\hat{c}_1$ , falls below  $c_1^*$ . Indeed, in this case, we can see from the above relationships that [\(8\)](#) is negative. To obtain the first-order condition – [\(8\)](#) equals zero – and given that the second-order condition

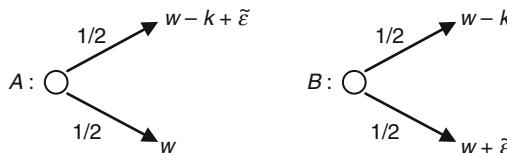
(negativity) is fulfilled by concavity of  $u - c_1$  must decrease, which yields  $\hat{c}_1 < c_1^*$ . Faced with future income risk, the DM reduces current consumption in order to build up precautionary savings.

While the link between prudence ( $u''' > 0$ ) and precautionary saving is instructive and elegant, one should nevertheless notice a discrepancy between the motivations for risk aversion ( $u'' < 0$ ) and for prudence ( $u''' > 0$ ). Risk aversion was defined as a general attitude toward risk, independently of a specific decision context in which this risk may arise, whereas prudence was developed in a specific context, the precautionary motive for saving. For this reason, prudence could appear as a concept which lacks generality.

Fortunately, more recent work, e.g., Eeckhoudt and Schlesinger (2006), has shown that prudence – like risk aversion – reflects a general attitude toward risk. To see this, consider the following simple binary lottery  $\mathcal{L}$ :



where  $w$  and  $k$  are positive amounts and  $\frac{1}{2}$  represents the probability of obtaining the stated amount of wealth. Assume now that a DM facing lottery  $\mathcal{L}$  must also face, with probability  $\frac{1}{2}$ , the prospect of a subsequent zero-mean risk  $\tilde{\varepsilon}$  that he or she does not like (because he or she is risk averse). If this DM adopts the principle that it is worthwhile “combining good with bad” (instead of ending with either “good-good” or “bad-bad”), then he or she will prefer to face the  $\tilde{\varepsilon}$  risk when his or her wealth is  $w$  instead of  $w - k$ . In other words, he or she will prefer  $B$  to  $A$  when faced with the choice between the two following lotteries:



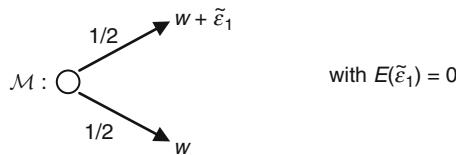
Put differently, he or she prefers “disaggregating pains.” Quite obviously, pains are disaggregated in  $B$ , not in  $A$ , and if  $B \succ A$  it is easily shown (see Eeckhoudt and Schlesinger 2006) that in the EU framework

$$B \succ A \Leftrightarrow u''' > 0$$

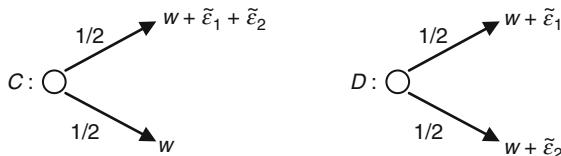
As a result, prudence has now a status equivalent to that of risk aversion: it reflects a very general attitude according to which risk-averse DMs prefer to avoid the concentration of pains. When faced with a prospect of a future risk (which they dislike), they feel less concerned by this prospect if it arises in a “good” state of nature – a state where the marginal utility of wealth is lower. Notice that this explanation of “prudence” provides the fundamental rationale for precautionary saving. In this case, the DM knows that he or she faces a future risk on his or her income. She prepares for this risk by saving more – which decreases his or her future marginal utility of wealth.

It is shown by Eeckhoudt and Schlesinger (2006) and by Eeckhoudt et al. (2009b) that this principle can be extended to higher orders of successive derivatives of  $u$ . Consider for instance the notion of *temperance* ( $u''' < 0$ ). Originally, temperance was linked to the following idea expressed by Kimball (1993): “It is reasonable to think that an unavoidable risk might lead an agent to reduce exposure to another risk even if the two risks are statistically independent.” Again, this definition is more related to a decision (“reduce exposure”) than to a fundamental preference. However, it can also be interpreted as a preference for disaggregating pains.

To see this, start from lottery  $\mathcal{M}$  where:



For a risk-averse DM,  $w + \tilde{\varepsilon}_1$  is the bad component of  $\mathcal{M}$  (it is dominated by  $w$ ). Now, suppose that with probability  $1/2$  this individual has to face another independent zero-mean risk  $\tilde{\varepsilon}_2$  (i.e., an additional pain) that he or she prefers to allocate to the good state. He or she prefers  $D$  to  $C$  where:



Eeckhoudt and Schlesinger (2006) show that in the EU framework

$$D \succ C \Leftrightarrow u'''' < 0, \text{ i.e., temperance.}$$

As a result, the sign of the fourth derivative of the utility function can be interpreted, like the sign of the second and third derivatives, as reflecting an attitude toward risk.

The reasoning can easily be extended to higher orders so as to give an interpretation to the alternating signs of successive derivatives of the utility function. When  $u$  exhibits such a property, shared by many (if not all) commonly used utility functions, the individual is said to be “mixed risk averse” (Caballé and Pomansky 1996).

Let us close this section with some general observations:

- While our discussion here concentrated on one-dimensional utility functions, the preference for “combining good with bad” also gives interesting insights about bi- or multidimensional utility: see Richard (1975), Eeckhoudt et al. (2007), and Tsetlin and Winkler (2009);
- Because the principle of “combining good with bad” is easily understood, it lends itself easily to experimentation: see Deck and Schlesinger (2010) and Ebert and Wiesen (2009);
- There is a natural link between the preferences discussed in this section and attitudes toward successive moments of a probability distribution of outcomes (a topic often

discussed in the finance literature): see Menezes et al. (1980), Chiu (2005), Roger (2010) or Ebert (2010);

- The principle of “combining good with bad” is also at the heart of the theory of asset prices. Prices of risky assets are obtained by discounting future payoffs. A fundamental principle arising from theoretical analyses, such as Arrow–Debreu pricing, stochastic discount factor analysis, or martingale pricing, is that prices of assets that pay in bad states of nature are “inflated,” compared to prices of assets that pay in good states. The individuals are prepared to pay more to “combine good with bad” (payoffs in a bad state), than to “combine good with good” (payoffs in a good state). The analysis shows that pricing occurs as if the market participants were adjusting the probability of states of nature, according to their overall market status (good or bad states). Given the above considerations, it is not surprising to find that prudence plays a role in defining a dividing line between good states and bad states for the pricing process (see Danthine and Donaldson 2005).

## Relative Risk Attitudes

---

So far, we have mostly paid attention either to the signs of the derivatives of the utility function, which indicate directions of preferences, or to the coefficient of absolute risk aversion  $A$  and its properties. This coefficient turns out to play a central role in the economics of risk when risks are additive.

In this section, we consider in more detail multiplicative risks from which one obtains inter alias the coefficient of relative risk aversion  $R$  defined in section [❸ Risk Aversion](#). Very early, three observations were made about this coefficient. First, from its definition (see section [❸ Risk Aversion](#)), this coefficient is equal to the elasticity of marginal utility of wealth with respect to current wealth. Secondly, many comparative statics results depend on the impact of wealth on the value of  $R$  and this impact is not unambiguously positive, negative, or null. Thirdly, it turns out also that many comparative results depend on a comparison between the value of  $R$  and unity. Notice indeed that because  $R$  is an elasticity, it has no dimension (in contrast to the risk premium  $\pi$  or to the coefficient of absolute risk aversion  $A$ ). For this reason, it can be compared to a pure number.

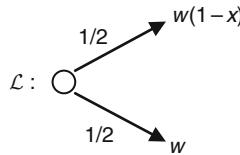
Note also that attention was paid more recently to the coefficients of absolute and relative prudence. Following the lead of Pratt (1964) in defining  $A$ , the coefficient of absolute prudence  $P_a$  was defined in Kimball (1990) as:

$$P_a = - \frac{u'''(E(\tilde{w}))}{u''(E(\tilde{w}))}$$

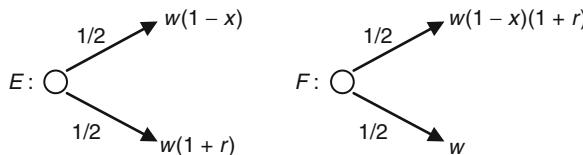
This coefficient is positive for a prudent, risk-averse DM. From there, the coefficient of relative prudence is defined as  $P_r = wP_a$ . While unity appears to be a sort of benchmark value for  $R$ , it is shown in Eeckhoudt and Schlesinger (2008) and in Danthine and Donaldson (2005) that the benchmark value for  $P_r$  is two.

In this section, we try to provide an intuitive interpretation for the benchmark value of unity applied to  $R$ . The reader can then refer to more technical papers that extend the reasoning to  $P_r$ , e.g., Eeckhoudt et al. (2009a).

Let a DM face a risk that, with probability  $\frac{1}{2}$ , reduces his or her wealth  $w$  to  $w(1 - x)$  where  $x$  is a positive number strictly smaller than unity. Hence, the DM faces a lottery  $\mathcal{L}$  represented as follows:



Now, this decision maker has the opportunity to benefit from a contingent increase in wealth in proportion  $r$ , where  $r$  is a strictly positive number. Depending on where he or she chooses to apportion this contingent increase, he or she faces one of the two lotteries  $E$  and  $F$ :



Which lottery will he or she choose? From section [Higher-Order Risk Attitudes: Prudence and Temperance](#), and the preference for combining good with bad, one tends to consider that  $F$  is preferred. However here, because of the multiplicative nature of the loss and gain, a new feature appears: lottery  $E$  has a higher mathematical expectation than lottery  $F$ , so that not all risk averters will prefer  $F$  to  $E$ . Only those who are “sufficiently” risk averse will prefer  $F$ .

What does it mean to be “sufficiently” risk averse? It can be shown that this is determined by whether or not  $R$  exceeds unity. To gain intuition about this result, consider a DM with a logarithmic utility function  $u(w) = \ln w$ . For this function, as is well known,  $R$  is constant and equal to 1. Besides, in this particular case,

$$E[\ln(\tilde{w})]_E = \ln w + \frac{1}{2} \ln(1 - x) + \frac{1}{2} \ln(1 + r) = E[\ln(\tilde{w})]_F$$

Consequently, with logarithmic utility, hence  $R$  constant and equal to unity, the DM is indifferent between  $E$  and  $F$ . To prefer  $F$ , the DM must be more risk averse than an individual with logarithmic utility. Using simple properties of the concave transformation of a utility function, it can be shown that this corresponds to  $R > 1$ . Conversely, a DM with  $R < 1$  will put relatively more emphasis on the fact that  $E$  has a larger mathematical expectation of final wealth than  $F$ , and he or she will prefer  $E$ .

To go to higher orders, beginning with relative prudence, one replaces  $(1 - x)$ , which corresponds to a known multiplicative loss, by  $(1 + \tilde{\varepsilon})$ , where  $\tilde{\varepsilon}$  is a zero-mean random variable (also a pain for a risk averter). Following basically the same argument as above, it can be shown that, to yield a preference for “combining good with bad,”  $P$ , must exceed 2. Details about this result are provided in Eeckhoudt et al. (2009a).

## Further Research: Applications – Insurance as an Example

The economics of risk has found numerous applications. All areas of economic life are concerned by questions of risk and uncertainty, but some of them more directly and more fundamentally than others. Health economics is one of them. This is due to the fact that, for any individual, the health status may change stochastically in time. However, health economics raises specific difficulties. The health status enters the utility function, besides wealth. Hence, the appropriate utility function is multidimensional and for this reason cross-derivatives of the utility function play a role in the analysis. For example, it is important for analytical purposes to have information on whether the marginal utility of wealth increases or decreases when the health status deteriorates. Concepts such as prudence and temperance become more complex to use when utility is multidimensional (see Eeckhoudt et al. 2007). Issues raised by health economics go much beyond the domain of the present survey. They have been addressed in numerous studies. Culyer and Newhouse (2000) edited a two-volume handbook of research on this topic.

Finance is another topic where the economics of risk has proved highly fruitful. Because finance deals with the transfer of purchasing power in time – from the present to the future and from the future to the present – and because the future is uncertain, finance is directly and fundamentally concerned by risk. In this case, the application of the economics of risk is straightforward. Wealth is appropriately considered as the one and only argument in the utility function and concepts such as risk aversion and prudence have direct applications in the study of financial behavior. Financial securities are easily conceived as lottery tickets providing specific payoffs in different states of nature. Indeed, as indicated before (see section ➤ [The Allocation of Risks Among Individuals](#)) any financial asset may be conceived as a portfolio of Arrow–Debreu contingent securities and its trading serves to transfer risks among individuals and institutions, reducing or eliminating the risks for some of them, and not necessarily increasing the risks for the others if a diversification process takes place through the transfer. The finance literature is huge and has witnessed a tremendous development since the 1960s with the applications of the economics of risk to financial economics issues. It is not appropriate to try to give even a brief account of this literature in this survey. The reader is rather referred to a textbook, such as by Danthine and Donaldson (2005). But within the broad area of finance, a specific aspect of risk transfer deserves a few paragraphs. This is insurance.

Insurance contracts are financial assets: they promise payments in the future, in specific states of nature, against the payment of a sure amount today (the insurance premium). As any other financial security, but still more obviously, they also represent bundles of Arrow–Debreu securities. They do not pay anything if a specific risk (the insured event) does not materialize. But they pay an indemnity in the states of nature where this event occurs. If the event represents a risk that is diversifiable in society (at least to some extent) and if problems of asymmetric information between insurers and insureds are not too large (see below), insurance companies will supply such contracts and will transform the risk into a predictable cost, to a large extent.

In the insurance domain, the demand for insurance represents a very fruitful application of the economics of risk aversion. Friedman and Savage (1948) had already shown that risk aversion is a necessary condition to produce a demand for insurance as soon as the insurance premium is “loaded,” i.e., more expensive than the actuarial value of the insured loss. A loading of the insurance premium is often observed and rationalized as a mean either to

cover insurance costs or to provide a profit to the insurer, or both. In fact, a loading may also be seen as a consequence of a lack of competition among insurers since the financial surplus provided by the loading – if any – supplements the financial income provided by the invested “technical” reserves of the insurer. If the latter income is high enough, competition among insurers should drive down the loading to a negative value.

The seminal paper on this topic was produced by Mossin (1968), who obtained two main results: (1) partial insurance coverage is optimal for an expected utility maximizer when the insurance premium incorporates a proportional loading (notice that this result was also present in Arrow (1963) and Smith (1968)) and (2) the demand for insurance decreases when wealth increases if the amount at risk remains fixed and the insured has decreasing absolute risk aversion (DARA). The second result is not controversial as long as the two required hypotheses are clearly indicated. But the first result seems to contradict casual observations. Most of the time, the insureds seem to prefer full insurance, although the premium is loaded and a proportional loading is common practice. With a lump sum loading, full insurance is optimal provided the loading does not exceed a certain threshold above which the individual prefers not to insure at all. If partial insurance is observed, this is often due to a limit being imposed on insurance coverage by the insurer, in his or her attempt to mitigate the effects of moral hazard and adverse selection (see below), not to a preference for partial insurance on the insured’s side.

Different explanations have been provided for Mossin’s result in numerous research papers (see Loubergé 2000, and Rees and Wambach 2008 for extended surveys of insurance economics; and Schlesinger 2000, for a specific presentation of the demand for insurance). The most interesting explanation is the recognition that Mossin’s analysis is based on the implicit assumption that the individual faces only one risk, whereas in reality, insureds are confronted with several risks having various characteristics: some of these risks are independent, others not; some of them are uninsurable, others may be insured, partially or fully; and still others have to be insured fully at a regulated price, due to social insurance provisions. Hence, given that the objective is to maximize the expected utility of final wealth, which is affected jointly by these various risks, the demand for insurance should be analyzed in a portfolio setting, not in isolation. Indeed, for given DM’s preferences and given insurance pricing, the decision to insure a given risk partially, fully, or not at all will depend on three elements:

- The insurable risk being considered
- The other risks faced by the DM, their relationship with the first risk, and their joint influence on final wealth
- The constraints imposed on risk transfers by the fact that some risks are not insurable (incomplete markets for insurance) and others are subject to mandatory insurance

The portfolio approach to insurance has proved particularly fruitful in this context. This portfolio approach was initially developed to study the optimal management of an insurer or reinsurer, seen as a financial intermediary (see Kahane and Nye 1975; Kahane 1977; Loubergé 1983). The portfolio approach was then used to reconsider models “à la Mossin” by Doherty and Schlesinger (1983), Mayers and Smith (1983), Turnbull (1983), Doherty (1984), von Schulenburg (1986), and Gollier and Scarmure (1994), among others. More recently, Rey (2003) considered the case where the “background” uninsurable risk is a health risk. The main tenet of this approach is that a risk should not be considered and managed per se, independently of the other risks faced by the individual or the institution. When investing

funds in the financial market, a risk-averse DM will take into account the imperfect correlations between the payoffs of the different traded assets. He or she will choose an optimal diversified portfolio that maximizes the expected utility of his or her final wealth and reflects his or her degree of risk tolerance (see Markowitz 1959). Similarly, in the insurance domain, the extent of coverage chosen for a given insurable risk will take into account the other risks and their joint influence on the DM's overall situation. For example, it may be optimal for him or her to insure a risk fully, even if the premium is loaded, due to a positive correlation between this risk and other risks for which no insurance is available. By insuring the insurable risk fully, although it would have been optimal to insure this risk partially if considered in isolation, the DM insures indirectly the uninsurable risk (see Doherty and Schlesinger 1983). A nice example is credit risk insurance for a small entrepreneur. His or her final wealth depends on the demand for his or her services that varies with the overall economic activity. This business risk is uninsurable. The entrepreneur faces another risk which may be insured, the risk that his or her customers do not pay when payments become due (credit risk). The two risks are positively correlated and it may thus be optimal for the entrepreneur to purchase full insurance for the credit risk even if the premium incorporates a proportional loading. Using the concept of prudence, Eeckhoudt and Kimball (1992) have also shown that the insurance purchasing decision is not independent of the other uninsurable risks faced by the DM, even if these risks are stochastically independent of the insurable risk.

These improvements in the theory of insurance demand closely paralleled similar advances in the theory of risk aversion under multiple sources of risk, e.g., Kihlstrom et al. (1981); Ross (1981) and Doherty et al. (1987). They were then followed by many papers examining comparative statics properties of insurance demand under "background risk" such as Meyer (1992), Dionne and Gollier (1992), Eeckhoudt et al. (1991, 1996), Gollier and Pratt (1996), Tibiletti (1995), Meyer and Ormiston (1996), Guiso and Jappelli (1998), and Meyer and Meyer (1998).

This literature fits well into the premises that have been chosen for the present survey – concentrate on the behavior toward risk observed for an individual decision maker. However, it should be stressed that insurance economics covers also other fruitful applications of the economics of risk to insurance problems: in particular, the study of insurance and reinsurance companies operations when risks cannot be fully diversified away, and the functioning of insurance markets when asymmetric information prevails between insurers and insureds. Issues raised by these considerations were reviewed in the handbook edited by Dionne (2000), and more recently in a series of papers published in *The Journal of Risk and Insurance* (September 2009) and in a survey study by Rees and Wambach (2008). These insurance issues go much beyond the optimal behavior of a risk-averse DM. They address societal issues, such as how should the economy deal with large risks having macroeconomic consequences, for instance changes in human longevity, large-scale catastrophes, and climate change.

A final set of issues concerns the problem of asymmetric information. Asymmetric information among market participants produces two kinds of imperfections in the market mechanism: moral hazard and adverse selection. Moral hazard occurs when the behavior of an agent (e.g., an insured) may have an impact on the outcome of a contract between this agent and a principal (e.g., an insurer), but the principal cannot observe whether the outcome is only due to the occurrence of an exogenous state of nature or whether the agent's behavior has played a role. Adverse selection occurs when the agent's characteristics have a statistical impact on the outcome of the contract, but the principal is unable to observe these characteristics. In both cases, the market mechanism is biased and the principal internalizes this bias in his or her

decisions to participate (or not) to the exchange, and to which extent. In the limit, the market does not exist (markets are not complete), or it works with second best arrangements, such as partial risk transfers or signaling by the agent that he or she is endowed with the “good” characteristics. Insurance markets are primary examples of markets where moral hazard and adverse selection are predominant, but financial markets and labor markets also offer good examples. What are the economic consequences of asymmetric information among market participants in the markets for risk transfers? If some markets work imperfectly or even do not exist for reasons of moral hazard and adverse selection, does this contradict the results derived from the Arrow–Debreu model of competitive allocation of risks among individuals? Should the public authorities intervene? How and with which expected consequences? To be able to address these questions with an appropriate economic toolkit, more advances in the economics of risk will be useful.

## Conclusion

---

While many topics have been left aside (as indicated in the introduction) we have covered in this survey two fundamental aspects of the economics of risk.

In the first sections we have reviewed well-known results, while in sections [Higher-Order Risk Attitudes: Prudence and Temperance](#) and [Relative Risk Attitudes](#) more recent developments about fundamental concepts have been analyzed. We have restricted these developments to models of one-dimensional utility, although, as already mentioned in section [Higher-Order Risk Attitudes: Prudence and Temperance](#), these developments start having implications for models of choices in a joint context of risk and multivariate arguments of utility (for instance, health and wealth).

The economics of risk has had a deep influence in many fields of economic theory. In the earlier years of its development (the 1960s and 1970s), it appeared as a specialized topic in microeconomic theory – the theory dealing with the behavior of economic agents and their interaction in markets. Nowadays, problems of risk, uncertainty, and information have been pervasive in all areas of microeconomics and finance. They have also started to have a tremendous impact at the macroeconomic level, as witnessed by developments in global finance and banking, or by the challenges of “sustainable development.” For these reasons, the concepts presented in the first sections of this survey are now part of the fundamental education of any economist. Given the limitations of this work, we have been unable to devote much space to applications, but we have at least provided a flavor of the direct applications that the economics of risk has found in insurance. The reference list below will hopefully help the interested reader to complete the brief account provided by this survey.

## References

---

- Allais M (1953) Le comportement de l’Homme rationnel devant le risque: critique des postulats et axiomes de l’Ecole Américaine. *Econometrica* 21:503–546
- Arrow KJ (1953) Le rôle des valeurs boursières pour la répartition la meilleure des risques. In: *Econométrie*, CNRS, Paris, pp 41–47. English version: The role of securities in the optimal allocation of risk-bearing. *Rev Econ Stud* 31:91–96
- Arrow KJ (1963) Uncertainty and the welfare economics of medical care. *Am Econ Rev* 53:941–969

- Arrow KJ (1965) Theory of risk aversion. In: Arrow KJ (ed) *Aspects of the theory of risk-bearing*. Yrjö Jahnsson Säätiöö, Helsinki, Reprinted in Arrow (1971)
- Arrow KJ (1971) Essays in the theory of risk bearing. North Holland, Amsterdam
- Artzner P, Delbaen F, Eber JM, Heath D (1999) Coherent measures of risk. *Math Financ* 9:203–228
- Bernoulli D (1738) Specimen theoriae novae de mensura sortis. *Comment Acad Sci Petropolinae* 5:175–192, Translated as Exposition of a new theory on the measurement of risk. *Econometrica* 22:22–36
- Bernstein P (1996) Against the gods. *The Remarkable Story of Risk*. Wiley, New York
- Black J, Bulkley G (1989) A ratio criterion for signing the effects of an increase in uncertainty. *Int Econ Rev* 30:119–130
- Borch K (1960) The safety loading of reinsurance premiums. *Skand Aktuarie Tidskr* 43:163–184
- Borch K (1962) Equilibrium in a reinsurance market. *Econometrica* 30:424–444
- Borch K (1990) Economics of insurance. North Holland, Amsterdam, Completed by Aase K, and Sandmo A
- Caballé J, Pomansky A (1996) Mixed risk aversion. *J Econ Theory* 71:485–513
- Chiu H (2005) Skewness preference, risk aversion and the precedence relations on stochastic changes. *Manage Sci* 51:1816–1828
- Crouhy M, Galai D, Mark N (2001) Risk management. McGraw Hill, New York
- Culyer AJ, Newhouse JP (2000) Handbook of health economics. Elsevier, Amsterdam, 2 vol
- Danthine JP, Donaldson JB (2005) Intermediate financial theory, 2nd edn. Academic, New York
- Debreu G (1953) Une économie de l'incertain. *Électricité de France*, Paris (Unpublished)
- Debreu G (1959) Theory of value. An axiomatic analysis of economic equilibrium. Wiley, New York
- Deck C, Schlesinger H (2010) Exploring higher order risk effects. *Rev Econ Stud* (forthcoming)
- Dionne G (ed) (2000) Handbook of insurance. Kluwer, Boston
- Dionne G, Gollier C (1992) Comparative statics under multiple sources of risk with applications to insurance demand. *Geneva Pap Risk Insur Theory* 17:21–33
- Dionne G, Eeckhoudt L, Gollier C (1993) Increases in risk and linear payoffs. *Int Econ Rev* 34:309–319
- Doherty N (1984) Portfolio efficient insurance buying strategies. *J Risk Insur* 51:205–224
- Doherty N, Schlesinger H (1983) Optimal insurance in incomplete markets. *J Polit Econ* 91:1045–1054
- Doherty N, Loubergé H, Schlesinger H (1987) Additive and multiplicative risk premiums with multiple sources of risk. *Scand Actuarial J* 53:41–49
- Drèze J, Modigliani F (1972) Consumption decision under uncertainty. *J Econ Theory* 5:308–335
- Dumas B, Allaz B (1996) Financial Securities, Chapman & Hall
- Ebert S (2010) On higher-order risk preferences, skewness and diversification. Working paper, University of Bonn, Bonn
- Ebert S, Wiesen D (2009) A methodology to test for prudence and third-order preferences. Bonn Econ Discussion Paper No 21
- Eeckhoudt L, Gollier C (2000) The effects of changes in risk on risk taking: a survey. In: Dionne G (ed) *Handbook of insurance*. Academic, Boston, pp 118–130, Chapter 4
- Eeckhoudt L, Hansen P (1980) Minimum and maximum prices, uncertainty and the theory of the competitive firm. *Am Econ Rev* 70:1064–1068
- Eeckhoudt L, Kimball M (1992) Background risk, prudence, and the demand for insurance. In: Dionne G (ed) *Contributions to insurance economics*. Kluwer, Boston, pp 239–254
- Eeckhoudt L, Schlesinger H (2006) Putting risk in its proper place. *Am Econ Rev* 96:280–289
- Eeckhoudt L, Schlesinger H (2008) Changes in risk and the demand for saving. *J Monet Econ* 55:1329–1336
- Eeckhoudt L, Gollier C, Schlesinger H (1991) Increases in risk and deductible insurance. *J Econ Theory* 55:435–440
- Eeckhoudt L, Gollier C, Schlesinger H (1996) Changes in background risk and risk-taking behavior. *Econometrica* 64:683–689
- Eeckhoudt L, Rey B, Schlesinger H (2007) A good sign for multivariate risk-taking. *Manage Sci* 53:117–124
- Eeckhoudt L, Etner J, Schroyen F (2009a) The value of relative risk aversion and prudence: a context-free interpretation. *Math Soc Sci* 58:1–7
- Eeckhoudt L, Schlesinger H, Tsetlin I (2009b) Apportioning of risks via stochastic dominance. *J Econ Theory* 144:994–1003
- Ekern S (1980) Increasing  $n$ th degree risk. *Econ Lett* 6:329–333
- Fishburn P (1989) Retrospective on the utility theory of von Neumann and Morgenstern. *J Risk Uncertain* 2:127–158
- Fishburn P, Porter B (1976) Optimal portfolios with one safe and one risky asset: effects of changes in rate of return and risk. *Manage Sci* 22:1064–1073
- Föllmer H, Schied A (2002) Stochastic finance: an introduction in discrete time. De Gruyter, Amsterdam
- Friedman M, Savage LJ (1948) The utility analysis of choices involving risk. *J Polit Econ* 56:279–304
- Gollier C (1995) The comparative statics of changes in risk revisited. *J Econ Theory* 66:522–536

- Gollier C, Pratt J (1996) Risk vulnerability and the tempering effect of background risk. *Econometrica* 64:1109–1123
- Gollier C, Scarmure P (1994) The spillover effect of compulsory insurance. *Geneva Pap Risk Insur Theory* 19:23–34
- Guiso L, Jappelli T (1998) Background uncertainty and the demand for insurance against insurable risks. *Geneva Pap Risk Insur Theory* 23:7–27
- Hadar J, Russell W (1969) Rules for ordering uncertain prospects. *Am Econ Rev* 59:25–34
- Hanoch G, Levy H (1969) The efficiency analysis of choices involving risk. *Rev Econ Stud* 36:335–346
- Kahane Y (1977) Capital adequacy and the regulation of financial intermediaries. *J Bank Finance* 1:207–218
- Kahane Y, Nye D (1975) A portfolio approach to the property-liability insurance industry. *J Risk Insur* 42:579–598
- Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. *Econometrica* 47:263–291
- Kihlstrom RE, Romer D, Williams S (1981) Risk aversion with random initial wealth. *Econometrica* 49:911–920
- Kimball M (1990) Precautionary saving in the small and in the large. *Econometrica* 58:53–73
- Kimball M (1993) Standard risk aversion. *Econometrica* 61:589–611
- Leland H (1968) Saving and uncertainty: the precautionary demand for saving. *Q J Econ* 82:465–473
- Levy H (1992) Stochastic dominance and expected utility: survey and analysis. *Manage Sci* 38:555–593
- Levy H (1994) Absolute and relative risk aversion: an experimental study. *J Risk Uncertain* 8:289–307
- Louberté H (1983) A portfolio model of international reinsurance operations. *J Risk Insur* 50:44–60
- Louberté H (2000) Development in risk and insurance economics: the past 25 years. In: Dionne G (ed) *Handbook of insurance*. Academic, Boston, pp 3–33, Chapter 1
- Markowitz HM (1959) Portfolio selection – efficient diversification of investments. Wiley, New York
- Mayers D, Smith CW (1983) The interdependence of individual portfolio decisions and the demand for insurance. *J Polit Econ* 91:304–311
- Menezes CF, Wang XH (2005) Increasing outer risk. *J Math Econ* 41:875–886
- Menezes C, Geissl C, Tressler J (1980) Increasing downside risk. *Am Econ Rev* 70:921–932
- Meyer J (1992) Beneficial changes in random variables under multiple sources of risk and their comparative statics. *Geneva Pap Risk Insur Theory* 17:7–19
- Meyer D, Meyer J (1998) Changes in background risk and the demand for insurance. *Geneva Pap Risk Insur Theory* 23:29–40
- Meyer D, Meyer J (2005) Relative risk aversion: what do we know? *J Risk Uncertain* 31:243–262
- Meyer J, Ormiston M (1985) Strong increases in risk and their comparative statics. *Int Econ Rev* 26:425–437
- Meyer J, Ormiston M (1996) Demand for insurance in a portfolio setting. *Geneva Pap Risk Insur Theory* 20:203–212
- Mossin J (1968) Aspects of rational insurance purchasing. *J Polit Econ* 79:553–568
- Pareto V (1896) *Cours d'Economie politique*. Rouge, Lausanne
- Pratt J (1964) Risk aversion in the small and in the large. *Econometrica* 32:122–136
- Rees R, Wambach A (2008) The microeconomics of insurance. *Found Trends Microecon* 4(1–2):1–163
- Rey B (2003) A note on optimal insurance in the presence of a non-pecuniary background risk. *Theory Decis* 74:73–83
- Richard S (1975) Multivariate risk aversion, utility independence and separable utility functions. *Manage Sci* 22:12–21
- Roger P (2010) Mixed risk aversion and preference for disaggregation: a story of moments. *Theory Decis* (forthcoming)
- Ross S (1981) Some stronger measures of risk aversion in the small and in the large with applications. *Econometrica* 49:621–638
- Rothschild M, Stiglitz J (1970) Increasing risk: I. A definition. *J Econ Theory* 2:225–243
- Rothschild M, Stiglitz J (1971) Increasing risk: II. Its economic consequences. *J Econ Theory* 3:66–84
- Sandmo A (1970) The effect of uncertainty on saving decisions. *Rev Econ Stud* 37:353–360
- Sandmo A (1971) On the theory of the competitive firm under price uncertainty. *Am Econ Rev* 61:65–73
- Savage LJ (1954) Foundations of statistics. Wiley, New York
- Schlesinger H (2000) The theory of insurance demand. In: Dionne G (ed) *Handbook of insurance*. Academic, Boston, pp 131–152, Chapter 5
- Smith V (1968) Optimal insurance coverage. *J Polit Econ* 79:68–77
- Tibiletti L (1995) Beneficial changes in random variables via copulas: an application to insurance. *Geneva Pap Risk Insur Theory* 20:191–202
- Tsetlin I, Winkler R (2009) Multiattribute utility satisfying a preference for combining good with bad. *Manage Sci* 55:1942–1952
- Turnbull S (1983) Additional aspects of rational insurance purchasing. *J Bus* 56:217–229

- Tversky A, Kahneman D (1992) Advances in prospect theory: cumulative representation of uncertainty. *J Risk Uncertain* 5:297–323
- von Neumann J, Morgenstern O (1947) Theory of games and economic behavior, 2nd edn. Princeton University Press, Princeton
- von Schulenburg M (1986) Optimal insurance purchasing in the presence of compulsory insurance and insurable risks. *Geneva Pap Risk Insur* 38:5–16
- Walras L (1874) *Eléments d'Economie Politique Pure* [Translated as Elements of Pure Economics] (trans: Corbaz L). Irwin, Homewood



# 6 Interpretation of Forensic Evidence

Reinoud D. Stoel · Marjan Sjerps

Netherlands Forensic Institute, The Hague, The Netherlands

<i>Introduction</i> .....	136
<i>Interpretation of Forensic Evidence by Means of the Likelihood Ratio</i> .....	
An HIV Test .....	139
Bayes' Rule .....	140
A DNA Match .....	141
How Likely Is It That the Suspect Is the Donor of the DNA Material Found at the Scene of Crime? .....	142
Fallacies .....	143
Non-DNA Forensic Evidence .....	144
Nonnumerical Likelihood Ratios .....	145
Numerical Likelihood Ratios Based on Continuous Data .....	147
<i>Special Issues</i> .....	148
The Prior Odds .....	148
Choosing the Hypotheses .....	148
Combining Evidence .....	150
Uncertainty .....	152
Context Effects and Bias .....	154
<i>Further Research</i> .....	156

**Abstract:** One of the central questions in a legal trial is whether the suspect did or did not commit the crime. It will be apparent that absolute certainty cannot be attained. Because there is always a certain degree of uncertainty when interpreting the evidence, none of the evidence rules out all hypotheses except one. The central question should therefore be formulated in terms of probability. For instance, how probable is it that the suspect is the offender, given the situation and a number of inherent uncertain pieces of evidence? The answer to this question requires the estimation, and subsequent combination, of all relevant probabilities, and cannot be provided by the forensic expert. What the forensic expert can provide is just a piece of the puzzle: an estimate of the evidential value of her investigation. This evidential value is based on estimates of the probabilities of the evidence given at least two prespecified hypotheses. These probabilities can subsequently be used by the legal decision maker in order to determine an answer to the question above, but they are, of course, not sufficient. They need to be combined with all the other information in the case. A probabilistic framework to do this is the Likelihood Ratio approach for the interpretation of forensic evidence. In this chapter we will describe this framework.

## Introduction

---

Imagine yourself being stopped by the police in a random traffic alcohol control. You did not drink a drop of alcohol, and you obtain a negative result on the breath analysis. However, you see the driver of the car in front of you taken into custody because of a positive test result on the breath analysis. You have read a recent popular scientific journal article in which the properties of breath analysis were discussed, and you remember that the test being used is correct in more or less 99% of the time. So, if the breath alcohol concentration is above the legal limit, there is a 99% chance of a positive test result, and, if the breath alcohol concentration is below the limit, then there is a 99% chance of a negative test result. Could you give an estimate of the probability that the breath alcohol concentration of the driver in the car in front of you is indeed above the legal limit? Is it a high probability or a low one?

It is tempting to give an answer to this question, and this answer is likely to be quite a high probability that the driver's alcohol concentration is above the legal limit. The truth of the fact is, actually, that an answer to this question cannot be given based on the information provided. An important part of the information that is needed is missing, and this part consists of the relative frequency of drivers that truly have an alcohol concentration above the legal limit. Aspects like driving style and prior speeding fines could of course also contribute. What we can legitimately conclude from the given information is the "evidential value" (to use the forensic science terminology; see Robertson and Vignaux 1995; Aitken and Taroni 2004) of the breath analysis (outside the field of forensics the term "relative risk" is often used for a similar statistical quantity). The evidential value helps in estimating the relevant probability, but it is certainly not enough. In our example the evidential value turns out to be equal to 99 (the exact calculation will be described later in section [Bayes's Rule](#)), and implies that a positive test result is 99 times more likely if a person has a breath alcohol concentration above the limit, than when it is below the limit. In other words, one could say that the positive test result supports the hypothesis that the breath alcohol concentration above the limit compared to the hypothesis that it is below the limit.

To illustrate the fact that the evidential value is not enough for estimating the probability that a random driver with a positive result indeed has a breath alcohol concentration above the limit, have a look at the data presented in [Table 6.1](#). [Table 6.1](#) presents two hypothetical

**Table 6.1**

Breath analyses in two populations with, respectively, 1% and 15% drunk driving

Sample 1	Positive result (+)	Negative result (-)	Total	Sample 2	Positive result (+)	Negative result (-)	Total
<i>ndd</i>	99	9,801	9,900	<i>ndd</i>	85	8,415	8,500
<i>dd</i>	99	1	100	<i>dd</i>	1,485	15	1,500
Total	198	9,802	10,000	Total	1,570	8,430	10,000

situations in which the breath test has been applied to test random samples of 10,000 persons from two populations. The first sample is an exact representation of a population in which 1% of the drivers are drunk-driving (*dd*) and 99% are not drunk-driving (*ndd*), and the second sample of a population in which 15% of the drivers are drunk-driving. The breath analysis test that has been applied has the properties mentioned above, and gives a positive result (+) or a negative result (-).

For the first sample, it is relatively easy to calculate that 99 persons will obtain an incorrect positive result (1% of 9,900), and that an equal 99 persons will obtain a correct positive result (99% of 100). So, for all 198 persons obtaining a positive test result, 99 are correct. The probability of drunk-driving given a positive test result is thus estimated as  $Pr_{\text{sample1}}(dd|+)$  = 99/198 = 0.50. For the second sample, on the other hand, 85 persons will obtain an incorrect positive result (1% of 8,500), and that 1,485 persons will obtain a correct positive result (99% of 1,500). So, for all 1,570 persons obtaining a positive test result, 1,485 are correct. The probability of drunk-driving given a positive test result is thus estimated as  $Pr_{\text{sample2}}(dd|+)$  = 1,485/1,570 = 0.95.

Thus, we cannot provide an estimate of the probability solely based on the properties of the breath analysis test [i.e.,  $Pr(+ | dd)$  = 0.99 and  $Pr(+ | ndd)$  = 0.01]. We need information on the prevalence of drunk-driving in the relevant population. The lower the prevalence, the lower the probability of drunk-driving given a positive test result [ $Pr(dd|+)$ ]. If we do base the estimate of the relevant probability solely on the properties of the breath analysis test, the obvious estimate will be equal to  $Pr(dd|+)$  = 99%, based on the fact that the test is correct in 99% of the cases. We would, however, then implicitly assume that the prevalence of drunk-driving is 50%. In other words, one out of every two drivers is drunk-driving. Both situations (i.e., “very low probability of drunk-driving” and “equal prevalence”) are illustrated in **Table 6.2**. Sample 3 comes from a population with 0.1% drunk-driving, and Sample 4 comes from a population with 50% drunk-driving. It is, again, relatively easy to see that the probability of drunk-driving, given a positive test result is equal to  $Pr_{\text{sample3}}(dd|+)$  = 99/1,098 ≈ 0.09 in Sample 3, and equal to  $Pr_{\text{sample4}}(dd|+)$  = 4,950/5,000 = 0.99 in sample 4.

What we have tried to make clear with this simple example, using a breath analysis test and population data, is that the probability of drunk-driving for a certain person depends on more than just the test result. The same test result may result in different probabilities that the person was in fact drunk-driving.

So, how does all this relate to forensic evidence interpretation? Well, imagine a case where a trace is compared to a possible source, for example, a fiber is compared to a suspect’s sweater. The answer to the question “How probable is it that the trace originated from this particular source?” cannot be provided by the forensic expert, since the expert will not be aware of all

**Table 6.2****Breath analyses in two populations with, respectively, 0.1% and 50% drunk driving**

Sample 3	Positive result (+)	Negative result (-)	Total	Sample 4	Positive result (+)	Negative result (-)	Total
<i>ndd</i>	999	98,901	99,900	<i>ndd</i>	50	4,950	5,000
<i>dd</i>	99	1	100	<i>dd</i>	4,950	50	5,000
Total	1,098	98,902	10,0000	Total	5,000	5,000	10,000

relevant information in the case. What the forensic expert can provide is an estimate of the evidential value based on estimates of the probabilities of the evidence given at least two hypotheses: quantities that are comparable to  $Pr(+ | dd)$  and  $Pr(+ | ndd)$ . They need to be combined by the legal decision maker with other evidence in the case.

The truth of the fact is that statistics and probability theory is not the most favorite topic of the average lawyer, although the tide is turning (Robertson and Vignaux 1995; Kaye 2010). Journals like “Jurimetrics” and “Law, probability and risk,” books (Aitken and Taroni 2004; Buckleton et al. 2005; Saks and Koehler 2005; Finkelstein and Levin 1990; Gastwirth 2000; Redmayne 2001) and conferences (e.g., International Conference on Forensic Inference Statistics, see <http://www2.unil.ch/icfis/>) try to fill the gap between the legal decision maker (i.e., the lawyers) and (forensic) statisticians. Forensic statistics is a broad field of expertise, ranging from applications of data analysis methods to determine, for example, sample sizes, to the development of methods for the evaluation and interpretation of forensic evidence.

The quantitative nature of DNA evidence stimulated these developments. The so-called Likelihood Ratio approach (LR approach) for the interpretation of evidence has proven very successful for DNA evidence, and can be regarded as a bridge between statistics and criminal law. The LR approach provides a measure of the evidential value, and explicitly defines the roles of the expert and the legal decision maker. At present this approach is actually being applied in the interpretation of DNA evidence, automatic speaker recognition, and speed calculations for colliding cars, where quantitative evidence is often available. It is expected that, eventually, for other forensic evidence such as fingerprints, glass, and comparative studies like handwriting and signature analysis such explicit calculations will be possible. We will later see that the LR method can be useful for assessing evidential value, even if numbers are lacking.

The aim of this chapter is to describe the LR approach for forensic evidence interpretation, and to discuss the main issues that arise in applying the approach in practice. We will start with another illustrative example taken from medical sciences, and subsequently discuss forensic applications. Special issues, such as combination of evidence and the effect of domain-irrelevant context information, will be treated in the last section.

## Interpretation of Forensic Evidence by Means of the Likelihood Ratio

The Likelihood Ratio approach is making strong progress in the forensic sciences as a method to evaluate the value of evidence. This section will discuss the details of this method. But let us first take a look at another illustrative example, this time from the medical sciences.

## An HIV Test

Let us assume that a certain HIV test will always result in a positive result for someone who has been infected with the HIV virus (i.e., a true positive test result). Let us assume, furthermore, that for a small number of subjects who have not been infected, the result will also be “positive” (i.e., a false positive test result). Now, consider Anna taking this specific HIV test, and obtaining a positive test result. The question that we will be investigating is: “What is the probability that Anna is actually infected with HIV, given the fact that she obtained a positive test result?”

The doctor is taken as the expert here, and the expert’s finding is the positive HIV test of Anna. This result (i.e., the positive test) is exactly what is to be expected of the specific test if Anna actually has been infected with HIV. But, on the other hand, if Anna has not been infected with HIV, the test result may also be an example of a rare, but possible, false positive result. Luckily, most of the persons without HIV obtain a negative test result, but some do obtain an incorrect positive test result. Anna could be one of those persons. However, taken altogether, we can state that a positive test result is much more probable if Anna has been infected [ $Pr(+ | hiv) = 1$ ], than when she is not infected [ $Pr(+ | no\ hiv) = \text{small}$ ].

Of course, Anna did undergo the HIV test in order to find out whether or not she has been infected with HIV, so she is interested in the doctors’ diagnosis. The doctor’s diagnosis, obviously, does depend on the test result, but it is not the only relevant piece of information that she uses. There is more information about Anna available to the doctor, and the doctor was already aware of this information before she knew the test result. The fact is that Anna belongs to an HIV risk group, namely, that of drug users who have used contaminated needles. So, even before the test was carried out, the doctor could already estimate the probability of infection to be quite high, and at least much higher than in the case of Anna not belonging to a risk group. This so-called prior probability (i.e., the probability of HIV infection prior to conducting the test) is adjusted upward because of the finding that the test is positive. The result is called the posterior probability, and with respect to Anna, this is the probability Anna has been infected “after” the positive test result is revealed. Two factors should therefore play a role in the diagnosis of the doctor: the prior probability of Anna being infected with HIV, and the test result. Adjusting the prior probability on the basis of new information is expressed mathematically in the LR approach. Since the prior probability differs for different persons, the same test result may lead to different diagnoses for different persons (i.e., a different posterior probability).

Does this make sense? Yes it does! In general, if a number of hypotheses may explain the same observation, the probability that one of these hypotheses actually is true is not only determined by the fact that the result indeed has been observed. The prior probability of the hypotheses plays a role too. The positive test result, in Anna’s case, may be explained by an HIV infection but may also be a false positive. The probability that Anna actually has been infected is not only determined by the fact that the test is positive, but also depends on the risk factors for Anna, or, in other words, on the prior probability that Anna is HIV positive.

The LR approach formalizes this line of reasoning. Anna’s risk factors affect the probability she has been infected with HIV even before she has been tested. This prior probability may be high or low, but will be adjusted upward if the test turns out to be positive. How this actually works can be described by means of a relatively simple, but extremely important, mathematical formula: Bayes’ rule. Bayes’ rule may be of aid in many decision problems like those encountered in medicine, as well as forensics.

## Bayes' Rule

The Likelihood Ratio approach is a direct application of the well-known *Bayes' rule*. Bayes' rule (or Bayes' theorem) was developed in the eighteenth century by the English clergyman Reverend Thomas Bayes (Bayes and Price 1763). The variant we will be using considers the ratio of the probabilities of two hypotheses (i.e., the odds). If the ratio is, for instance, equal to ten to one (i.e., 10:1), the first hypothesis is ten times as probable as the second hypothesis. This ratio of probabilities of the hypotheses can be considered before certain findings (i.e., the evidence) are taken into account (prior odds) and after they have become known (posterior odds). Bayes' rule shows how, in this context, the ratio of the probabilities changes due to the findings:

Prior odds  $\times$  Likelihood Ratio = Posterior odds

in term of probabilities:

For ease of notation explicit mentioning of the background information  $I$  is omitted here (cf. Aitken and Taroni 2004). This information is assumed known in all probabilities.

In order to illustrate the use of Bayes' rule, let us use the breath analysis test data from Sample 4 described in the introduction. We concluded that, in this case, the probability of drunk-driving ( $dd$ ) given a positive test was equal to  $Pr_{\text{sample4}}(dd|+) = 0.99$ . The calculation of this probability can also be obtained by using Bayes' rule.

We define:

- Hypothesis 1: The driver was drunk-driving ( $dd$ )
  - Hypothesis 2: The driver was not drunk-driving ( $ndd$ )
  - Evidence: The driver has a positive result on the breath analysis test (+)

and we assume (corresponding to Sample 4 in [Table 6.2](#))

- $Pr_{\text{sample4}}(dd) = 0.5$
  - $Pr_{\text{sample4}}(ndd) = 0.5$
  - $Pr_{\text{sample4}}(+ | dd) = 0.99$
  - $Pr_{\text{sample4}}(+ | ndd) = 0.01$

The prior odds of Hypothesis 1 versus Hypothesis 2 are thus equal to Prior odds<sub>sample4</sub> = 0.5/0.5 = 1, and the LR is equal to 0.99/0.01 = 99, resulting in Posterior odds<sub>sample4</sub> = 1 × 99 = 99. So the odds of drunk-driving versus not drunk-driving, after a positive test result, are equal to 99:1 which corresponds to  $Pr_{\text{sample4}}(dd|+) = 99/(99 + 1) = 0.99$ .

With this in mind, it is easy to show what the posterior probability will be if we take different prior odds, for instance those of Sample 1. Prior odds<sub>sample1</sub> = 0.01/0.99 ≈ 0.0101, resulting in Posterior odds<sub>sample1</sub> = 0.0101 × 99 ≈ 1. So the odds of drunk-driving versus not drunk-driving, after a positive test result are equal to 1:1 which corresponds to the  $Pr_{\text{sample1}}(dd|+)$  = 1/(1 + 1) = 0.5 that we saw in the introduction. Again this illustrates that the same test result may lead to different posterior probabilities for different persons.

In the next section we argue that the forensic expert should report the LR (i.e., the evidential value), and that she should not report on the posterior probability. Let us focus on an example taken from forensic sciences: matching DNA profiles.

## A DNA Match

---

A trace containing DNA is found at the scene of a crime, which may belong to the offender. In order to facilitate calculation, this example will assume the DNA material to be of very poor quality, such that only a small part of the DNA profile can be visualized. This partial profile shows, among other things, that it concerns a man. The forensic DNA expert concludes that the suspect's DNA profile matches the partial profile from the crime scene trace. She determines that the probability that a randomly selected other man, unrelated to the suspect, will match the partial DNA profile found at the scene of crime [i.e., the “*random match probability*”] is equal to 0.001 (i.e., 0.1%). How probable is it that the DNA found at the crime scene originates from the suspect? Just as was the case in the breath analysis and HIV example above, the DNA expert cannot give a direct answer to this question. What the DNA expert will do is “translate” this question into two mutually exclusive, not necessarily exhaustive, hypotheses:

- Hypothesis 1: The suspect is the donor of the DNA material.
- Hypothesis 2: An unknown man, unrelated to the suspect, is the donor of the DNA material.

Subsequently, the expert will compute the Likelihood Ratio, defined as the ratio of the probability of a match given that the suspect is the donor of the DNA material (Hypothesis 1) versus the probability of a match given that an unknown man, unrelated to the suspect, is the donor of the DNA material. Based on the assumption that no mistakes have been made in the chain of custody from finding the trace up to reporting of the DNA match, the scientist will reason as follows:

- If the suspect is the donor of the trace, their profiles will match. The probability of the finding of the DNA analysis is therefore equal to 1 if Hypothesis 1 is true [i.e.,  $Pr(\text{match}|\text{suspect is donor}) = 1$ ].
- If an unknown other man is the donor of the DNA material, it would be fairly coincidental that this material matches the suspect's profile. The chances of this occurring are 1 in 1,000. The probability of the finding of the DNA analysis is therefore 0.001 if hypothesis 2 is true [ $Pr(\text{match}|\text{unkown man is donor}) = 0.001$ ].

The ratio of the probability of the findings given the two hypotheses (i.e., the LR) is therefore  $Pr(\text{match}|\text{suspect is donor})/Pr(\text{match}|\text{unkown man is donor}) = 1/0.001 = 1,000$ . In this simple example, it is clear that the LR is thus equal to 1/random match probability. This LR implies that the match is 1,000 times more likely if Hypothesis 1 is true than if Hypothesis 2 is true. The results (i.e., the match between the suspects DNA profile and the DNA profile of the crime scene trace) are thus much better explained by Hypothesis 1, than by Hypothesis 2. In other words, the result offers more support for the first hypothesis than for the second.

## How Likely Is It That the Suspect Is the Donor of the DNA Material Found at the Scene of Crime?

The issue is again that the answer depends, as it did in the HIV example, not just on the results of the DNA analysis, but also on the prior odds of the hypotheses. Or, in terms of Bayes' rule, the answer depends on the LR as well as the prior odds. The LR is determined by the rarity of the DNA profile, expressed in the random match probability. The prior odds are determined on the basis of other information in the case; this could be tactical information, or other forensic evidence. For example, suppose a crime took place aboard a container ship at full sea, and only ten other men can be the donor of the DNA in the trace. Moreover, assume all these men to equally qualify as donors of the DNA trace, but the only DNA profile available is that of the suspect. In this case, the odds in favor of the suspect being the donor before the DNA analysis is taken into consideration are 1 to 10. The prior odds of Hypothesis 1 versus 2 are therefore  $0.1$ . Bayes' rule ( $\text{prior odds} \times \text{LR} = \text{posterior odds}$ ) shows that the posterior odds are equal to  $0.1 \times 1,000 = 100$ . This means that, after the DNA analysis is taken into consideration, the odds are 100 to 1 in favor of the suspect being the donor of the trace material and not one of the other members of the crew. ➤ *Table 6.3* illustrates how the probability that the DNA belongs to the suspect depends on both the rarity of the DNA profile and the number of men aboard the ship, or, in other words, on both the LR and the prior odds. From this table it may be inferred that the rarer the DNA profile, the larger the LR, and, consequently, the stronger the DNA evidence against the suspect. Also, the fewer men aboard the ship, the larger the prior odds, and hence the larger the posterior odds.

In the HIV example, the doctor can make a statement on the posterior odds of HIV infection, because she has specialist knowledge on both the prior odds and the LR. The DNA

■ **Table 6.3**

**The table illustrates the effect of the DNA match on the odds of Hypothesis 1: “The suspect is the donor of the DNA material” versus Hypothesis 2: “One of the other men aboard the ship is the donor of the DNA material.” We consider four different prior odds, corresponding to, respectively, 1, 10, 100, and 1,000 other crewmembers. The final column is calculated from the posterior odds. The random match probability of the DNA profile of the trace is 1 in 1,000 versus 1 in 1 billion, corresponding to LRs of 1,000 and 1 billion**

Random match probability	LR	Prior odds	Posterior odds	Probability suspect is donor
1 in 1,000	1,000	1:1	1,000:1	0.999
		1:10	100:1	0.99
		1:100	10:1	0.909
		1:1,000	1:1	0.5
1 in 1,000,000,000	1,000,000,000	1:1	1,000,000,000:1	0.999999999
		1:10	100,000,000:1	0.99999999
		1:100	10,000,000:1	0.9999999
		1:1,000	1,000,000:1	0.999999

expert, on the other hand, only has specialist knowledge of the relative frequency of the DNA profile, i.e., of the LR. The DNA expert will usually not have specialist knowledge or a complete overview of the other information of the case, which determines the posterior odds. She will not wish to express an opinion in this respect. This means, in line with the previous examples, but in sharp contrast to common understanding, that a DNA expert does not calculate the probability that the trace originates from the suspect (or from someone other than the suspect). This posterior probability, as we have seen, depends on the prior odds. The DNA expert limits herself to a conclusion about the LR. As the LR is in the present case equal to 1/random match probability, she will, in practice, only report this probability.

## Fallacies

---

Suppose that the DNA profile of a suspect, Peter, matches with the partial DNA profile of a bloodstain found at a crime scene. The donor of the trace must be a man, and the random match probability equals 1 in 1,000. The prosecutor argues as follows:

- 1p The probability that a randomly chosen man matches this profile is 1 in 1,000.
- 2p The probability that someone else is the donor of the bloodstain is thus 1 in 1,000.
- 3p The probability that Peter made the bloodstain is thus close to 1 (i.e., 99.9%).

The lawyer in this case, on the other side, walks a different route and reasons:

- 1l Approximately 8 million men live in the Netherlands.
- 2l There are about 8,000 other men that match the partial DNA profile.
- 3l In a total of 8,001 men with this DNA profile, including Peter, the probability that Peter made the bloodstain is thus 1/8,001 (i.e., 0.012%).
- 4l The evidential value of the match is very small.

Who is right? Unfortunately neither of them, because both fell prey to notorious fallacies! The prosecutor made a fallacy that has been termed the prosecutor's fallacy (Thompson and Schuman 1987). The fallacious argument here is that the probability under Argument 1p is definitely not the same as the probability under Argument 2p. Argument 1p refers to the frequency of the partial profile in the relevant population of men (i.e.,  $P(\text{match}|\text{randomly chosen man is the donor})$ ), while Argument 2p is about the probability that the bloodstain is not Peter's, given the match (i.e.,  $P(\text{randomly chosen man is the donor}|\text{match})$ ). As we saw in the example of section ➤ A DNA Match this last probability is a posterior probability that not only depends on the frequency of the partial profile in the population but also on the prior probability that Peter is the donor of the bloodstain. The probability under Argument 2p cannot be based on just the random match probability.

The lawyer in the DNA case fell prey to a fallacy in the literature known as the defense attorney's fallacy (Thompson and Schuman 1987). Here the fallacious argument is that the matching partial DNA profile is of little value. The reasoning is however based on some implicit assumption. First, the defense lawyer assumes that the offender must be among the population of Dutch men, which is obviously a fairly arbitrary assumption about the group of potential offenders. Second, he assumes that each person within this population has an equal probability of being the offender. All men are equally likely the offender, regardless of their age, location, or

health. However, while these assumptions may be true, it is definitely fallacious to conclude that the evidence is of little value. Because, no matter how large the population of possible offenders, this population is decreased in number with a factor of 1,000 due to the matching DNA profile.

## Non-DNA Forensic Evidence

---

With its roots in DNA evidence the usefulness of the LR approach is fortunately not limited to this type of evidence. In general, the LR approach has been, and is, frequently applied for all types of forensic evidence where there is some kind of uncertainty affecting interpretation, such as shoe marks, fingerprints, signatures, bullets, and photographs. As was the case for DNA evidence, the experts start with a translation of the main question into a set of (at least) two mutually exclusive hypotheses. Ideally these hypotheses are given to the expert by the legal decision maker, prosecution or defense lawyer, or they follow seamlessly from the information provided in the case. Sometimes the specific hypotheses are the result of extensive communication between the lawyer and expert. While the forensic investigations are led by these hypotheses, the important point to consider is that each of these hypotheses already has some probability of being true attached to it, as per the previous examples. The ratio of these probabilities defines the prior odds of the hypotheses. It is based on other evidence and tactical information in the case, for instance, information that the suspect had a financial dispute with the victim. Thus, an estimate of this ratio falls outside the field of expertise of the forensic expert that is investigating the new forensic piece of evidence. What the forensic expert can do is to provide estimates of the probabilities of the evidence given each of the relevant hypotheses, just like for DNA evidence. The shoe print examiner, for instance, in seeing a match on a special feature between the footprint at a crime scene and a shoe of a suspect, could give his subjective estimate of the probability of finding this specific feature if a shoe, other than that from the suspect made the trace.

It is the task of the forensic expert to give her conclusion on the LR, that is, her conclusion on the ratio of probabilities of the evidence, given the hypotheses. Based on all the other evidence and information in the case, the legal decision maker could make an estimate of the prior odds of the hypotheses, and could subsequently use the LR provided by the expert to update the prior odds. As an example, if the evidence is equally probable to be obtained under both hypotheses the LR will have a value of 1. Multiplication of the prior odds by this LR does not affect the size of the prior odds. This is what we would call “neutral evidence,” and the posterior odds of the hypotheses will be equal to the prior odds of the hypotheses. Neutral evidence thus offers no support for either hypothesis. If, on the other hand, the evidence is much more probable given Hypothesis 1, than given Hypothesis 2, the LR will be much larger than 1. That is, if

$$Pr(\text{evidence}|\text{hypothesis 1}) > Pr(\text{evidence}|\text{hypothesis 2}),$$

than  $LR > 1$  and this would thus be evidence supporting Hypothesis 1 with respect to Hypothesis 2. The prior odds will consequently be multiplied by a number much larger than one, resulting in posterior odds that are much larger than the prior odds. The opposite is also true when the LR is much smaller than 1. The evidence would then be supporting Hypothesis 2.

So, in summary, the LR method takes the LR as a measure of evidential strength with respect to the hypotheses. The role of the forensic expert is limited to reporting the evidential strength to the legal decision maker. It is the role of the latter to assess the probabilities of the hypotheses, by updating her prior beliefs about the odds using the expert's LR. Hence, assessments of the probability of guilt, but also of the probability of forensic hypotheses such as the probability that a trace originated from a particular source, are made by the legal profession, not by the expert. Moreover, any decisions based on these assessments are outside the province of the forensic expert. Thus from a decision perspective, the forensic expert has a very different role than many other experts, like for instance a doctor.

## Nonnumerical Likelihood Ratios

---

In some fields the forensic scientist can actually compute LR based on hard data and a statistical model. The best example is forensic DNA analysis, but other fields are developing toward numerical LRs (e.g., fingerprint analyses, glass analysis). Many fields of expertise exist, however, where numerical LRs cannot yet, if ever, be computed because the relevant data are lacking, and the forensic expert cannot numerically compute the probabilities. The forensic expert provides a more or less subjective estimate of the probabilities based on her knowledge, expertise, and experience in the field. This estimate may be partially dependent on some data, and generally it will not be very precise. To distinguish between a LR based on solid data and statistics, and a LR based on subjective opinion, many forensic laboratories use a set of verbal qualifiers to report the latter LR. Assume, again, for example, that a forensic shoe print examiner compares a shoe mark found at the scene of a crime with the shoe of a suspect. The expert will, in such cases, consider at least two hypotheses, for example:

- Hypothesis 1: The suspect's shoe made the mark.
- Hypothesis 2: Another shoe with a similar sole pattern and size made the mark.

The similarities and differences identified between the shoe of the suspect and the shoe mark found at the crime scene constitute the evidence. The shoe print examiner will be able to (subjectively) estimate the probability of evidence, given that the suspect's shoe made the mark and given that another shoe with a similar sole pattern and size made the mark. The term probability is used in this context as a criterion for the extent to which the findings are "striking" or "fit" the hypotheses. If the evidence is more in line with Hypothesis 1 than with Hypothesis 2, it constitutes evidence in support of Hypothesis 1. A striking similarity, such as a long cut in the shoe sole whose shape and position correspond with a "line" in the shoe mark will be more in line with Hypothesis 1 than Hypothesis 2. The better the findings fit with Hypothesis 1 relative to Hypothesis 2, the stronger the evidence will be. In the interpretation of her findings, the scientist will therefore establish how probable it would be to observe the findings if Hypothesis 1 is true, and when Hypothesis 2 is true. In such cases, she could conclude, for example, that:

- "The findings of the investigation are much more probable if the mark was made by the shoe of the suspect than if the mark was made by another shoe with a similar sole pattern and size."

It may also be the case that the findings are more probable in the event that the second hypothesis is true, for example if the shoe was secured immediately after the crime was

committed, but the long cut cannot be “found” in the shoe mark. The evidence is then supporting Hypotheses 2. For example:

- ▶ “The findings of the investigation are much more probable if another shoe with a similar sole pattern and size made the mark than if the shoe of the suspect made the mark.”

Another example is a case in which a forensic biometrical expert examines some vague CCTV images of a person robbing a bank. The observed similarities and differences of certain biometrical features of the body and the face between a suspect and the images of the perpetrator are taken here as the evidence. The relevant hypotheses may be:

- Hypothesis 1: The suspect is the perpetrator.
- Hypothesis 2: Another person, not related to the suspect, is the perpetrator.

The expert subsequently compares the probability of the evidence (i.e., the observed similarities and differences) under the two hypotheses and concludes, for instance, “the evidence is much more probable if Hypothesis 1 is true than if Hypothesis 2 is true,” or, in other words, “the evidence is much more probable if the suspect is the perpetrator, than if another person, not related to the suspect, is the perpetrator.” This constitutes a verbal approach to conclude that the numerator of the LR is much larger than the denominator of the LR, and implies that the LR is very large. The expert bases her conclusion only partly on numerical data, because the data available may be a very small, not representative set of the relevant population. If the expert observes a butterfly tattoo on the right cheek of both the perpetrator and the suspect, for instance, she may give a subjective estimate of the rareness of this feature, without being able to give a numerical estimate (e.g., 1 out of every 50,000 adult men has a butterfly-like tattoo on the right cheek). This last estimate would require a population study of tattoos.

It is very important to note that the conclusion (A) “the evidence is much more probable if Hypothesis 1 is true than if Hypothesis 2 is true” does not imply conclusion (B) that it is very probable that Hypothesis 1 is true, or any other probability statement about hypotheses since this would also require prior probabilities. The prior probabilities are not part of conclusion (A). Whatever the prior odds are, conclusion (A) implies that the posterior odds will be higher than the prior odds.

Interpretation of such verbal conclusions appears to be quite difficult, and the forensic service providers using the LR approach spend much time and effort teaching and explanation such conclusions. There is, however, quite a broad consensus among forensic scientists that the LR approach is the preferred approach for presenting forensic evidential strength, and that posterior odds and posterior probabilities are outside the province of the expert. This consensus has not yet been reached among legal practitioners, as illustrated by several UK cases (*R v Doheny and Adams [1997] 1 Crim App R 369*, and recent cases like *R v T [2010]*). The fact that verbal LRs are difficult to interpret should, however, not be so surprising. People just have great difficulties in interpreting (ratios of) conditional probabilities, and this is possibly worsened if (ratios of) conditional probabilities are expressed verbally. Beyond that, much has been written on the statistical and mathematical skills of the legal decision maker, which hardly have had any education in statistics during their education.

So, it may not be clearly defined what exactly is the definition of a verbal conclusion, and these definitions may not be constant for all fields of expertise. Some authors have argued for using some sort of numerical scale corresponding to the verbal conclusions

(see Lucy 2005; Evet 1998). Forensic laboratories in Ireland and the UK use such a scale (AFSP 2009). Sensitivity tables, such as presented by Robertson and Vignaux (1995), that show the effect of a certain LR on various values of the prior odds could also aid in interpretation.

## Numerical Likelihood Ratios Based on Continuous Data

Up till now we have only considered nonnumerical LRs, as well as numerical LRs data based on discrete data, for which we could compute the probabilities. However, quite often evidence is obtained from measurements on continuous data. For instance, suppose that we are comparing glass fragments. One glass fragment is found in the clothes of a person suspected of smashing a window and subsequent burglary. The other six fragments are reference samples taken from the broken window by the police. Suppose that we are comparing the refractive index of these fragments to the refractive index of the glass fragment found in the clothes. If we measure each fragment once, we are then comparing six measurements of the reference samples to a single measurement of the fragment found in the clothes, on a continuous scale.

The form in which the evidential value is determined for continuous data is similar to that for discrete data (Aitken and Taroni 2004). Since continuous measurements are being considered of both the crime scene material and the suspect material, respectively  $y$  and  $x$ , the probabilities of the evidence given the relevant hypotheses are replaced by probability density functions  $f$ . In the glass example,  $y$  is a vector of six measurements, and  $x$  is a single number. The LR now becomes, in general,

$$\frac{f(\text{evidence}|\text{Hypothesis 1})}{f(\text{evidence}|\text{Hypothesis 2})} = \frac{f(x, y|\text{Hypothesis 1})}{f(x, y|\text{Hypothesis 2})}$$

which is after mathematical simplification (not presented here), and assumption of independence of  $x$  and  $y$  given Hypothesis 2 (so-called conditional independence assumption), equal to

$$\frac{f(y|x, \text{Hypothesis 1})}{f(y|\text{Hypothesis 2})}.$$

The numerator of this LR is a so-called predictive distribution that considers the distribution of the crime scene measurements, conditional on Hypothesis 1 being true and the measurements on the suspect material. The denominator is a marginal distribution, which considers the distribution of the measurements on the crime scene material conditional on Hypothesis 2. In evaluating such a LR two sources of variation need to be considered, the within source variation (the numerator), and the between source variation (the denominator). With respect to the glass fragment, in the numerator it is considered how well the refractive index of the fragment fits within that of the window. Inhomogeneity of the refractive index within the glass window and measurement uncertainty play a role. In the denominator, it is considered how well the refractive index of the fragment fits other possible sources of the fragment. The rarity of the refractive index of the fragment plays an important role. Instead of a ratio of two probabilities, the LR now consists of a ratio of two probability densities. We refer to Aitken and Taroni (2004) for a detailed introduction to the appropriate approaches for Likelihood Ratios for continuous data, univariate as well as multivariate. Multivariate observations result for instance from analyzing the elemental composition of the glass.

## Special Issues

---

### The Prior Odds

---

The Likelihood Ratio approach can be thought of as a sequential information processing system for determining the probabilities of the corresponding hypotheses. The ratio of these probabilities, the odds, changes each time a piece of information is added, as specified by Bayes' rule. The prior odds, based on the tactical information and other evidence in the case, are multiplied by the LR of the evidence under consideration. If, subsequently, a new piece of evidence is evaluated, the "old" posterior odds become the new prior odds. This process is called updating, and goes on until all evidence has been evaluated. With every piece of evidence we consider the odds of the hypotheses before processing this information (the prior odds), and after (the posterior odds). The posterior odds after the first piece of evidence become the prior odds of the next piece of evidence. This process requires that the hypotheses that are considered for each piece of evidence are exactly the same. It is important to note that "true" prior odds do not exist, because they depend on which evidence is evaluated first. Similarly, the posterior odds and the LR itself are relative terms: They depend on the information assumed to be known at the time of information processing. The information may include legal evidence, but also more general, tactical, information such as the location of the crime.

In practice it is usually impossible for the legal decision maker to determine the prior odds precisely. However a rough distinction between large and small may well be possible. Consider, for example, a case concerning a suspect gun, where the question is whether a bullet from the crime scene was shot with this gun, or with another firearm. The hypotheses could be formulated as:

- Hypothesis 1: The bullet was shot with the suspect's gun.
- Hypothesis 2: The bullet was shot with another firearm.

It will be clear that the prior odds of Hypothesis 1 versus 2 are much larger in the situation that gun and bullet are both found on the crime scene, than when the bullet is found at the crime scene, and the gun in a safe deposit on the other side of the world. It has been argued that the Bayesian framework for interpreting evidence is useless because the starting point, the prior odds, is unclear. It is certainly true that it is difficult to deal with this if one actually wants to use the framework for calculations. However, one can turn the argument around and argue that legal decision making is a difficult process by nature. The Bayesian framework makes problems such as the existence of prior odds transparent. It is important to realize that using another reasoning framework cannot make the problem disappear. When one is interested in the posterior odds, like a legal decision maker, dealing with prior odds cannot be avoided. That the Bayesian framework makes this problem visible is an advantage, not a disadvantage. Hence, dealing with prior odds is a problem that is inherent to legal decision making, and any framework of reasoning that does not make this clear is useless.

### Choosing the Hypotheses

---

A hypothesis can be seen as a possible scenario that offers an explanation of a given event in a criminal lawsuit. As will have become clear from the above, the hypotheses play a crucial role

in the determination of the evidential value of the evidence under consideration, and the explicit formulation of the hypotheses is critical. Since the task of the expert is to provide (an estimate) of the probability of the evidence given the hypotheses, small changes in the formulation of the hypotheses will result in different LRs and consequently in a different evidential value. Often, especially in the more complex cases, hypotheses are determined through extensive interaction between the expert(s) and the legal decision maker. Given the fact that in the majority of cases at least one of the hypotheses reflects the scenario that is in favor of the suspect(s), the suspect and his defense lawyer should play a role here. The legal decision maker is aware of all information in the case, and did come, in the first place, with a specific question to the expert. The expert on the other hand is aware of all relevant scientific issues that need to be taken into account in formulating the hypotheses. In order to review the most relevant issues in formulating hypotheses, we will now discuss the hypotheses for the examples from prior sections (i.e., DNA and the shoe print).

Let us start with the simple example concerning a small biological trace where the forensic DNA expert found the suspect's DNA profile to match the partial profile from the crime scene. In such DNA cases the hypotheses are often straightforward, though this need not always be the case. The hypotheses we have formulated before are on the so-called source level. Source-level hypotheses give possible explanations to the corresponding evidence by stating who, or what, the source is of the trace that has been found. In this case the first hypothesis is easy to formulate since one of the questions the legal decision maker wants to answer is whether the suspect is the donor of the trace; therefore, the first hypothesis will be:

- Hypothesis 1s: The suspect is the donor of the DNA material.

The second hypothesis then provides an alternative explanation on the source level that is not accounted for by the first hypothesis. Since the trace did contain gender information, this is included in the alternative hypothesis that is formulated as:

- Hypothesis 2s: An unknown man, unrelated to the suspect, is the donor of the DNA material.

So, if the suspect is not the donor of the trace, it must have been another man, unrelated to the suspect, who left the trace. These hypotheses thus only provide information as to who is the donor of the crime scene trace. If the question was not about the source of the trace but whether some activity led to the trace the hypotheses may be formulated as, for instance:

- Hypothesis 1a: The suspect has been in close physical contact with the victim.
- Hypothesis 2a: The suspect has never been near the victim.

And, if the question was about the offense itself, the hypotheses may be:

- Hypothesis 1o: The suspect physically abused the victim.
- Hypothesis 2o: An unknown man, unrelated to the suspect, physically abused the victim.

Hypotheses can thus be formulated at different broad levels, the so-called hierarchy of propositions (Cook et al. 1998). The first level is the source level (s), the second level is the activity level (a), and the third level is offense level (o). As we climb higher on the hierarchy, the hypotheses more closely reflect the offense that is considered by the legal decision maker, and more factors come into play. On the source level of the DNA case the forensic expert will determine the random match probability of the crime scene trace in the relevant population.

At this level he may also consider issues like relatedness or laboratory errors. However, the fact that somebody is the donor of a trace does not necessarily imply that he or she performed some specific kind of activity, or crime. The trace may be there as a consequence of numerous types of activities, many of them being unrelated to the crime. Besides requiring measurement, observation, and analysis of information about the occurrence of the evidence in the relevant population, hypotheses on the activity level require judgments on how the trace ended up at the crime scene. In general, hypotheses on the activity level require more information than hypotheses on the source level. Usually the activity level requires estimates of the probability that traces are transferred during the activities considered. Moreover, estimates are required of the probability that these traces persist on the receptor and subsequently recovered by the police or forensic examiner. The resulting LR analysis usually becomes intractable when all this information is incorporated. Bayesian networks have been proposed for analyzing such complex structures (Taroni et al. 2006).

Often many hypotheses can be formulated in a given case. In forensic DNA comparison of a mixed DNA profile from a crime scene with the DNA profiles from two suspects A and B, for instance, possible hypotheses could be:

- The DNA mixture consists of DNA from A and B.
- The DNA mixture consists of DNA from A and an unknown person unrelated to A or B.
- The DNA mixture consists of DNA from B and an unknown person unrelated to A or B.
- The DNA mixture consists of DNA from two unknown persons not related to A or B.
- The DNA mixture consists of DNA from A and the brother of B.

Because it is often impossible to consider all hypotheses that could make sense in a specific case, and a large number of such hypotheses lead to an illegible report, the expert limits the number of hypotheses, and usually chooses just two hypotheses. The exact formulation of hypotheses is important, and subtle differences in wording can have a major effect on the resulting LR. Another restriction is that scientific hypotheses should be mutually exclusive, which implies that if one of the hypotheses is correct, the other(s) should be incorrect. They need not be exhaustive (contrary to what is stated in AFSP 2009). Finally, legal constraints prevent the use of hypotheses regarding the guilt of the suspect. In summary, the hypotheses should be formulated according to the following rules:

- At least two hypotheses should be considered.
- Each hypothesis needs to be relevant for at least one of the parties involved.
- The LR depends on the exact formulation of hypotheses.
- The hypotheses are mutually exclusive, but not necessarily exhaustive.
- The hypotheses do not explicitly consider the guilt of the suspect.
- Careful consideration should be made regarding the alternative hypothesis, since it explicitly defines the relevant population (cf. the “reference class problem,” Colyvan and Regan 2007).

## Combining Evidence

In the beginning of this section, the Likelihood Ratio approach was described as a sequential information processing system in which the prior odds can be combined with multiple pieces of evidence. Combining evidence is indeed an important aspect of the Likelihood Ratio

approach. The Likelihood Ratio enables the combined evaluation of multiple pieces of evidence in an intuitive and simple way. It is helpful to split the discussion of combining evidence into two factors: (1) conditionally independent versus dependent evidence and (2) same hypotheses versus different hypotheses. For the ease of presentation, we will focus on the combination of two pieces of evidence,  $E_1$  and  $E_2$ , but extension to multiple pieces of evidence is straightforward.

The easiest way of combining evidence is when the hypotheses are the same and the evidence is conditionally independent. This is, for example, the case if a DNA profile is obtained from a finger mark, and the evidence is a matching DNA profile ( $E_1$ ), and a matching fingerprint ( $E_2$ ) with hypotheses:

- $H_1$ : The suspect is the donor of the fingerprint.
- $H_2$ : An unknown man, unrelated to the suspect, is the donor of the fingerprint.

Since knowing a person's DNA profile used for forensics purposes does not tell us much about his fingerprint, we may treat them as independent pieces of information. If we make the crucial assumption that the DNA and the fingerprint are from the same donor, the subsequent LR is simply the product of the two separate LRs:

$$LR_{E_1, E_2} = \frac{P(E_1, E_2 | H_1)}{P(E_1, E_2 | H_2)} = \frac{P(E_1 | H_1)}{P(E_1 | H_2)} \times \frac{P(E_2 | H_1)}{P(E_2 | H_2)} = LR_{E_1} \cdot LR_{E_2}$$

If the evidence is not (conditionally) independent the LR becomes more complicated:

$$LR_{E_1, E_2} = \frac{P(E_1, E_2 | H_1)}{P(E_1, E_2 | H_2)} = \frac{P(E_1 | H_1)}{P(E_1 | H_2)} \cdot \frac{P(E_2 | E_1, H_1)}{P(E_2 | E_1, H_2)} = LR_{E_1} \cdot LR_{E_2 | E_1}$$

On the other hand, if the evidence is considered to be conditionally independent, but the hypotheses are not the same, the evidence cannot be combined statistically unless one formulates hypotheses on a different level in the hierarchy. If the evidence in a burglary case consists of glass fragments on a hammer of a suspect that match the broken window ( $E_1$ ), and paint fragments on the broken window match the paint of the hammer ( $E_2$ ), the evidence cannot be combined on the source level. Indeed, the hypotheses would then be formulated as:

- $H_{1,E1}$ : The glass fragments originate from the broken window at the crime scene
- $H_{2,E1}$ : The glass fragments originate from another float glass object and
- $H_{1,E2}$ : The paint fragments originate from the suspect's hammer
- $H_{2,E2}$ : The paint fragments originate from another object

However, the evidence can be combined on the activity level that does not only address the source of the fragments, but also the activity that was performed during the crime. In this case hypotheses on the activity level could be formulated as:

- $H_{1a}$ : The hammer of the suspect smashed the crime scene window.
- $H_{2a}$ : Another object smashed the crime scene window.

The evidence (i.e., matching glass and paint fragments) can now be combined, but the resulting LR is a qualitatively different one than on the source level, since it addresses the LR of a different set of hypotheses. However, the activity-level hypotheses may more closely correspond to the level at which the evidence needs to be evaluated in court. In evaluating

evidence on the activity level, it will be clear that more factors come into play and need to be taken into account by the forensic experts (see Aitken and Taroni 2004; Taroni et al. 2006). Bayesian networks are currently viewed as the method of choice to derive the evidential value in these types of cases.

An interesting question from a legal and philosophical perspective may be whether the expert is allowed at all to combine evidence, especially if part of this evidence falls outside his field of expertise. This occurs, for instance, in cases where the pathologist bases his conclusions on the results of the toxicologist. In another case in the Netherlands, for example, involving a man that was accused of injuring a boy by driving over him, a fiber expert, a chemist, and an expert specialized in shape comparison combined their conclusions and jointly signed the report. The defense lawyer in this case raised the question of whether this was allowed. Both the court and the higher court accepted the report, however.

## Uncertainty

---

The likelihood ratios presented above are based on ratios of two probabilities, the values of which are usually estimated using samples taken from the relevant populations. We focus here on discrete data. For continuous data the focus should be on estimation of the probability densities, and the corresponding confidence interval will be more complex. Nevertheless, a natural consequence of sampling is sampling error, which results in the estimates not being exactly equal to their population values. As a consequence, the LR estimated by a forensic expert does contain some uncertainty. The estimated LR is thus more, or less, precise depending on the size of the samples. As described by James Curran regarding DNA evidence

► ...The pertinent question is "how different could the frequencies (i.e., probabilities) be and what effect will this have on the LR?" In order to answer this question it is necessary to try and quantify the uncertainty about the allele frequencies. This uncertainty is often expressed in statistical literature as a confidence interval.... The Bayesian counterparts are the credible interval...Unfortunately, measures of sampling error are not commonly presented in court (Curran 2005, p 116).

Much work needs to be done on modeling, and presenting, uncertainty of LRs for forensic evidence, especially for continuous data. The main issue can be illustrated effectively with the following example. Consider a case in which the police encounter an echo in a questioned audio fragment, and suspects the two callers to be inside the same car while calling, presumably for giving each other an alibi. The police subsequently sent the audio fragment to the forensic laboratory in order to obtain an estimate of the strength of the evidence (i.e., the echo in the audio fragment [E]) concerning the question did caller and receiver share the same vehicle? Together with all parties involved the following set of hypotheses was formulated:

- $H_1$ : Caller and receiver of the phone call were inside the same car.
- $H_2$ : Caller and receiver of the phone call were not inside the same car.

In such a rare case with a considerable lack of relevant data, the forensic expert working on the case could decide to perform an experiment in which audio fragments of phone calls are recorded with exactly the same mobile phone under two conditions. Here, the experiment consists of 41 phone calls recorded under the first condition (i.e., caller and receiver inside the

same car), with the result that in 40 of these calls an echo was present. In the other condition 43 phone calls were recorded (i.e., caller and receiver not inside the same car), and in just one of these calls an echo was present. An obvious estimate of the probability of the presence of an echo under each of the two conditions is obtained by:

$$P(E|H_1) = 40/40 + 1 = 40/41 \approx 0.976$$

$$P(E|H_2) = 1/1 + 42 = 1/43 \approx 0.023$$

One immediately sees that the presence of an echo is much more likely if caller and receiver are inside the same car ( $H_1$ ), than if they are not inside the same car ( $H_2$ ). A straightforward measure of the strength of the evidence is subsequently obtained by the LR, being equal to the ratio of the probabilities of the evidence under the two hypotheses as described above:

$$LR = P(E|H_1)/P(E|H_2) = 0.976/0.023 \approx 42,$$

The probabilities are estimated by means of relatively small sample sizes. Although the estimates of  $P(E|H_1) \approx 0.976$  and  $P(E|H_2) \approx 0.023$  are the current best estimates, the true, but unknown, values of probabilities may be smaller or greater than the obtained estimates due to sampling fluctuation. The true likelihood ratio may consequently be smaller, or greater, than the value of 42. To illustrate the relationship between uncertainty and sample size, one may think of the same probabilities [i.e.,  $P(E|H_1) \approx 0.976$  and  $P(E|H_2) \approx 0.023$ ] being the result of much larger sample sizes, for instance, 4,100 and = 4,300. We intuitively expect (and indeed can prove) that the uncertainty will be smaller.

Assessing the uncertainty in LRs has received quite some attention in the statistical literature (Brown et al. 2001; Dann and Koch 2005; Koopman 1984) and attention in the forensic literature is increasing (e.g., Curran 2005; Curran et al. 2002; Morrison 2010). Several methods have been proposed and, from a statistical perspective, it is not always easy to choose the best approach. The choice should depend, among other things, on the required coverage (and length), on the population proportions, on the obtained sample size, and perhaps on ones preference for a “frequentist” of “Bayesian” estimation procedure. Without going into detail, a 95% (frequentist) confidence interval based on the Score method results, for the results of the experiments mentioned above, in a lower bound of the LR of 8, and an upper bound of 237. A 95% (Bayesian) credible interval using a Jeffrey’s prior results in a lower bound of the LR of 9, and an upper bound of 382.

There has been a fundamental discussion among forensic statisticians about the meaning of uncertainty of the LR. Some have adopted the view that the calculated LR should not be considered as an estimate of some “true” but unknown LR. Consequently, it is argued that there is no sense in studying the uncertainty of the LR. We believe there are strong arguments for the notion of a “true” but unknown value of the LR, given the relevant hypotheses and background information, and that it is important to consider the uncertainty. Ignoring the uncertainty can be strongly misleading. The question remains, however, which method should be used in a given case, and what the forensic expert should report. Should she report the best estimate (i.e.,  $LR = 42$ ), should she report the interval, both, or should she only report the lower or upper bound? We refer again to James Curran

- ▶ ‘...demonstrating that these errors have been calculated is the only correct and convincing response to the almost predictable questions about the size of the database and the validity of the results. The most appropriate comment on the subject perhaps comes from Buckleton

(2004) "An analyst who is prepared for such a cross examination will definitely present better evidence to the court than one who chooses to answer, 'would it make a difference?'" (Curran 2005, p 116).

Please note that we have addressed here mainly the issue of uncertainty due to sampling error, assuming the correct populations have been defined in the hypotheses. Other factors exist that result in uncertainty. For instance, uncertainty about the appropriate relevant population, or reference class, may also be important to take into account. Colyvan and Regan (2007) note with respect to this:

- ▶ 'Uncertainty will always be prevalent in legal decisions. Mathematical models of evidence provide some relief from uncertainty but can suffer from reference-class problems. This introduces a source of meta-uncertainty, which may well be impossible to eliminate. Decision theory, however, is in the business of recommending the best course of action in the face of uncertainty and thus can provide a valuable framework for approaching legal decisions and their associated uncertainties.'

## Context Effects and Bias

---

An important but slightly different aspect from the aspects of the interpretation of forensic evidence we touch upon in this chapter is that of context effects and cognitive bias. Cognitive processes play an important role in human action and consequently also in expert judgment. Brain and cognitive processes enable us to prioritize and group large amounts of information, and draw conclusions, despite possibly ambiguous and incomplete information (see Dror 2009, for an excellent discussion in the forensic context). They allow us, for instance, to distinguish our next door neighbor from a burglar in the dark, or to cross a busy street. These processes, that form the basis of human intelligence and expertise, however, also lead to specific vulnerabilities. Human intelligence and, in particular, expertise include aspects of mental representations that may also result in selectivity and biased information processing.

The capabilities and vulnerabilities expressed by Dror are inherent to human cognition, and thus two sides of same coin. Eliminating one cannot be done without sacrificing the other. What we can do in the forensic context is to check the vulnerabilities and minimize their effects by proper training and the development of appropriate methods and procedures. The vulnerabilities we will focus on here are so-called cognitive biases. These are the human tendencies to draw conclusions in certain situations based on cognitive factors rather than on observations or evidence. For instance, a firearm examiner can be biased by knowing that large amounts of drugs were found at the suspect's house. The term often used in the literature for designating this bias, is the term "observer effect." In forensic science, the term context effect is frequently used to indicate that the results of a forensic investigation can be partly determined by the circumstances in which the research is conducted, and especially by (irrelevant) case information known to the forensic expert (Thompson 2009). Frequently in the forensic literature, the term "confirmation bias" is used as a synonym. The interested reader is referred to Saks et al. (2003), and Risinger et al. (2002), for a more thorough treatment of the topic and for more references from the psychological literature. In this section we will shortly reflect on the methods that can be used in practice to reduce the effect of the context on forensic expert judgment. Such

methods may especially be important for forensic disciplines that rely on more subjective judgments for which no quantitative LR can be computed based on representative data.

The question that arises is of course whether these effects do actually occur in the practice of forensic evidence evaluation. Dror and Cole (2010) give an overview of studies of context effects in the forensic sciences. The few studies that have been carried out by forensic scientists and behavioral scientists partially contradict each other. Hall and Player (2008), for example, come to the conclusion that it is not such a big issue. Other studies, however, show a clear effect. Two studies are worth discussing. First, William Thompson (2009) shows, in an ad hoc experiment, that knowledge of the DNA profile of the suspect by forensic DNA experts may affect the interpretation of the disputed (partial and complex) DNA profile. By focusing on the DNA profile of the suspect it may occur that not all information is properly taken into account. Second, the well-designed experiment by Dror and colleagues (Dror et al. 2006) provides evidence for context effects in forensic fingerprint experts. Five experts were given a finger mark and fingerprint that they had already previously evaluated in a real case as a match (i.e., identification). This time, however, fingermark and fingerprint were presented in a context strongly pointing toward a non-match. Four of the five experts who participated in his experiment concluded that finger mark and print did not match, and they thus drew a different conclusion than they had previously drawn based on exactly the same mark and print. One might argue, of course, that these are two fairly limited experiments, and that several studies contradict each other. On the other hand, one might also argue that if even the smallest chance of bias exists it would be best for all of the forensic examinations to be performed in such a way that the effect of bias is minimized.

The context information that can result in bias may be of several different forms: (1) “base rate” information, (2) domain irrelevant case information, and (3) information obtained from the reference material. The first is information that is, in principle, independent of the particular case and the specific investigation, but which can have an effect on the expectations of the researcher. Due to the work of the police investigators, most evidence presented for forensic evaluation is incriminating, thereby making it in advance already tend toward a certain conclusion, and then unconsciously looking for information that supports the expected conclusion. In fact, this is one of the arguments used in the discussion whether judges in the Netherlands should specialize in criminal law or whether they should work in different areas of the law: it is feared that seeing a large number of cases in which the suspect usually is clearly guilty will cloud the judgment.

Domain irrelevant case information is all information that is not explicitly necessary for the expert performing the research. A confession, for instance, is almost always irrelevant information for a forensic fingerprint examiner.

The third form of biasing information is inherent in the way investigation is conducted, and was described in the previous paragraph. If the evidence is analyzed simultaneously with the reference material, the interpretation of the relevant features of the evidence may become partly determined on what the expert has seen in the reference material. Thompson (2009) mentions with respect to this the “Texas Sharp Shooter fallacy” named after a legendary Texan

- ‘...who fired his rifle randomly into the side of a barn and then painted a target around each of the bullet holes. When the paint dried, he invited his neighbors to see what a great shot he was. The neighbors were impressed: they thought it was extremely improbable that the rifleman could

have hit every target dead centre unless he was indeed an extraordinary marksman, and they therefore declared the man to be the greatest sharpshooter in the state.'

The effect of context information can be reduced as much as possible by keeping such information away from the experts carrying out the investigation. Tactical information, reports and other descriptions of the circumstances of the case, and results of other investigations are then not known to the expert carrying out the investigation. Easier said than done, this could be accomplished by using a stepwise procedure, where the person doing the intake of the case and the coordination is not the same as the person performing the actual investigation. Furthermore, to prevent target shifting, the evidence should be interpreted as much as possible, before the expert looks at the reference material. In summary, the expert doing the forensic investigations should be as blind as possible to all irrelevant case information. The addition of fake cases, with opposite conclusions, to the case flow could be a possible solution to the effect of base rate information. Although effective in theory, it is clear that this is not very practical solution because of the difficulty to create realistic cases, with the experts actually not being aware of the corresponding case being a fake case. Forensic scientists frequently build up a collection of "scary cases," cases with unexpected results, and use this collection to train new colleagues. Such cases are also discussed at conferences and in the literature. This may induce an important "exemplar effect," which has never been to topic of detailed study.

In the literature, the evidence lineup (Risinger et al. 2002) is sometimes advocated to overcome context effects. An evidence lineup is comparable to the well know Oslo (eye-witness) confrontation approach. Despite the high face validity the usefulness of the evidence lineup is severely limited in practice due to the selection of the fillers (i.e., reference material). Many of the recommendations of the evidence lineup can be traced back to the authoritative publication of Risinger, Saks, Thompson, and Rosenthal in 2002, but already in 1984 and 1987 Miller published on its application to handwriting research. A thorough analysis of the evidence lineup seems however to learn that its usefulness in investigating criminal cases is relatively small. In contrast the others, Risinger et al. themselves seem to be well aware of the difficulties where they write

- ▶ 'Proper evidence lineups present some nontrivial problems of design, requiring the Evidence and Quality Control Officer both to determine what would constitute appropriately similar foil specimens and to arrange to obtain them. This process would obviously be easier for some types of examinations than for others. Unfortunately, it may often be most difficult precisely where it is most needed, in those areas, such as handwriting identification, with the least instrumentation and greatest subjectivity' (pp. 49–50).

---

## Further Research

In this chapter we have described the LR approach for the evaluation and interpretation of forensic evidence. It has become clear that the forensic experts should report on the LR, and by doing so they report on the probability of the evidence given the relevant hypotheses. The legal decision maker should use this information, and combine it with all other information in the case to come to a verdict. While the basics of the LR approach are clear and the definition of the LR is straightforward, determination and reporting in

practice is not without difficulties, and much research is needed. We will now shortly address the most important issues in future research.

First of all, research should focus on methods to compute quantitative LRs for all forensic disciplines. In DNA analysis, reporting quantitative LRs (or the random match probabilities) is rather standard. For other areas like fingerprint analysis and forensic glass comparison, among some others, it is expected that numerical LRs can be reported in the near future. For many other areas computing quantitative LRs will remain science fiction for quite a long time. It is expected, therefore, that the evidential value will remain to have a considerable amount of subjectivity, in the sense that the evidential values are not derived from theory and hard data. Purely objective science is an illusion in every area – however, we should aim for it. Research should focus on how best to define the relevant populations, how to collect the data, and how to compute LRs based on those data. Deriving appropriate formulas that include measures of uncertainty, proper sampling strategies, and protocols to avoid context effects is indispensable.

A recent development in forensic science is a decision analysis perspective (Taroni et al. 2010). One of the topics considered is that in general forensic scientists tend to prefer erring in favor of the suspect over erring in favor of the prosecution, and they rather underestimate the strength of the evidence than overstate it. Consequently, when reporting confidence intervals or credibility intervals one may prefer erring on one side rather than erring on the other side. This may result in asymmetric intervals around the point estimate. Furthermore, decisions about the research strategy (which items of evidence are investigated, which analyses are performed in these items and in what order) are interesting to study from a decision analysis perspective. Practical software for performing such analyses is the Bayesian network software.

Given the fact that often multiple pieces of evidence are investigated in a given case, combination of evidence will also need to be considered in future research. Bayesian networks (Taroni et al. 2006) could be of great aid, especially if independence is violated and if statistical combination can only be performed on the activity level, with possibly much more factors affecting the result than on the source level. Bayesian networks could also aid in the understanding of the conclusion by the legal decision maker. Communication of results and conclusions such that all scientific details are properly addressed and understood by the legal decision is one of the major future challenges for all scientist and practitioners working in forensics.

## References

- Aitken CGG, Taroni F (2004) Statistics and the evaluation of evidence for forensic scientists, 2nd edn. Wiley, Chichester
- Bayes T, Price R (1763) An essay towards solving a problem in the doctrine of chance. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, M. A. and F. R. S. Philos Trans R Soc Lond 53: 370–418
- Brown LD, Cai T, DasGupta A (2001) Interval estimation for a binomial proportion (with discussion). Stat Sci 16:101–133
- Buckleton J, Triggs CM, Walsh SJ (eds) (2005) Forensic DNA evidence interpretation. CRC, Boca Raton
- Colyvan M, Regan HM (2007) Legal decisions and the reference class problem. Int J Evidence Proof 11:274–285
- Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA (1998) A hierarchy of propositions: deciding which level to address in casework. Sci Justice 38: 231–239
- Curran JM (2005) An introduction to Bayesian credible intervals for sampling error in DNA profiles. Law Prob Risk 4:15–126
- Curran JM, Buckleton JS, Triggs CM, Weir BS (2002) Assessing uncertainty in DNA evidence caused by sampling effects. Sci Justice 42:29–37

- Dann RS, Koch GG (2005) Review and evaluation of methods for computing confidence intervals for the ratio of two proportions and considerations for non-inferiority clinical trials. *J Biopharm Stat* 15:85–107
- Dror IE (2009) How can Francis Bacon help forensic science? The four idols of human biases. *Jurimetrics* 50:93–110
- Dror IE, Cole S (2010) The vision in 'blind' justice: expert perception, judgment and visual cognition in forensic pattern recognition. *Psychon Bull Rev* 17:161–167
- Dror IE, Charlton D, Peron A (2006) Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Sci Int* 156:74–78
- Evet IW (1998) Toward a uniform framework for reporting opinions in forensic science case work. *Sci Justice* 38:198–202
- Finkelstein MO, Levin B (1990) Statistics for lawyers. Springer, New York
- Gastwirth JL (ed) (2000) Statistical science in the courtroom. Springer, New York
- Hall LJ, Player E (2008) Will the introduction of an emotional context affect fingerprint analysis and decision-making? *Forensic Sci Int* 181:36–39
- Kaye DH (2010) The double helix and the law of evidence. Harvard University Press, Cambridge
- Koopman PAR (1984) Confidence intervals for the ratio of two binomial proportions. *Biometrics* 40:513–517
- Lucy D (2005) Introduction to statistics for forensic scientists. Wiley, Chichester
- Morrison GS (2010) Forensic voice comparison. In: Freckleton I, Selby H (eds) Expert evidence. Thomson Reuters, Sydney
- Redmayne M (2001) Expert evidence and criminal justice. Oxford University Press, Oxford
- Risinger MD, Saks MJ, Thompson WC, Rosenthal R (2002) The Daubert/Kumbo implications of observer effects in forensic science: hidden problems of expectation and suggestion. *California Law Review* 90:1–56
- Robertson B, Vignaux GA (1995) Interpreting evidence: evaluating forensic science in the courtroom. Wiley, Chichester
- Saks MJ, Koehler JJ (2005) The coming paradigm shift in forensic identification science. *Science* 309:892–895
- Saks MJ, Risinger MD, Rosenthal R, Thompson WC (2003) Context effects in forensic science: a review and application of the science of science to crime laboratory practice in the United States. *Sci Justice* 43:77–90
- Taroni F, Aitken CGG, Garbolino P, Biedermann A (2006) Bayesian networks and probabilistic inference in forensic science. Wiley, Chichester
- Taroni F, Bozza S, Biedermann A, Garbolino P, Aitken CGG (2010) Data analysis in forensic science: a Bayesian decision perspective. Wiley, Chichester
- Thompson WC (2009) Painting the target around the matching profile: the Texas sharpshooter fallacy in forensic DNA interpretation. *Law Prob Risk* 8:257–276
- Thompson WC, Schuman EL (1987) Interpretation of statistical evidence in criminal trials—the prosecutor's fallacy and the defense attorney's fallacy. *Law Hum Behav* 11:167–187

# **7 Risks and Scientific Responsibilities in Nanotechnology**

*John Weckert*

Charles Sturt University, Canberra, ACT, Australia

<b><i>Introduction</i></b> .....	<b>160</b>
<b><i>Some Risks of Nanotechnology</i></b> .....	<b>160</b>
Nanoparticles .....	161
Grey Goo .....	162
Privacy .....	162
Cyborgs .....	163
Nanodivide .....	164
<b><i>Models of Science</i></b> .....	<b>164</b>
The Linear Model .....	164
The Social Model .....	165
<b><i>Values in Research</i></b> .....	<b>168</b>
<b><i>The Science/Ethics Interface</i></b> .....	<b>169</b>
Interface 1 .....	169
Interface 2 .....	169
Interface 3 .....	170
Interface 4 .....	170
<b><i>Moral Responsibility</i></b> .....	<b>171</b>
<b><i>Risk and Moral Responsibility in Nanotechnology</i></b> .....	<b>172</b>
Risk with Nanoparticles .....	172
Risk with Grey Goo .....	173
Privacy .....	174
Cyborgs .....	174
Nanodivide .....	175
<b><i>Conclusion</i></b> .....	<b>175</b>
<b><i>Further Research</i></b> .....	<b>175</b>

**Abstract:** This chapter outlines a number of risks of nanotechnology and considers whether scientists can be held responsible, and if so, to what extent. The five risks discussed are representative of different kinds of risks and the list is not comprehensive: nanoparticles, privacy, grey goo, cyborgs, and nanodivides. The extent to which scientists can be held responsible for harms resulting from their research depends on the nature of science and here two models are outlined and assessed; the linear model and the social. The relationship of moral values to scientific research is examined with respect to both models and four interfaces are considered: the issues of concern to ethics committees, moral values in the acceptance or rejection of hypotheses, setting research agendas, and scientific responsibility. This leads to a discussion of responsibility itself, and on the basis of this, the five risks noted at the beginning of the chapter are revisited and an assessment given of the moral responsibility of scientists in each case.

## Introduction

---

Science and technology are often credited with making our lives much better through improved health, labor saving devices, better transportation, and so on. On the other hand, science and technology are often blamed for many of the problems of the developed world: various health problems, pollution, environmental degradation, breakdown of communities, and weapons of mass destruction, to name a few. In other words, science and technology are responsible for much that is good but also much that is bad. But what can be said about the responsibilities of individual scientists or groups of scientists? Should they be held responsible for the harms that sometimes result from their work? This chapter will attempt to sort out some of the issues relevant to risk and scientific responsibility. It will first consider a number of types of risks associated with nanotechnology and then outline two models of science, each of which has different implications for responsibility. This leads into a more detailed look at responsibility before a return to nanotechnology risks.

## Some Risks of Nanotechnology

---

Not much can be done in life without taking risks. Living is risky. From the moment of conception, we are surrounded by risks, many of which could prove fatal. But most, luckily, do not eventuate and most of those that do leave us relatively unscathed. We still worry about them of course, and this is probably a good thing, within reason. Preventative action can be taken and many avoided. Science and technology in general are risky enterprises, with the twin problems of unintended consequences and dual use. Worries are particularly prevalent in new and emerging technologies, and nanotechnology is no exception (for convenience, nanotechnology is used here to include nanoscience as well). In this chapter, a number of different kinds of risks of nanotechnologies will be considered. These various kinds of risks highlight different aspects of risk, risk perception, risk management, and so on.

Some of the things discussed as risks here may not be considered as risks by all, but each will be justified as a potential risk in the relevant discussion. The risks pertain to nanoparticles, grey goo (defined later), privacy, cyborgs and nanodivides.

Before examining these, it is instructive to look at an overview of nanotechnology and risk from a recent report. Four generations of nanotechnology are distinguished with potential risks posed by each.

## First Generation

Passive (steady function) nanostructures, for example, nanostructured coatings and non-invasive; invasive diagnostics for rapid patient monitoring *From 2000 –*

*Potential risk:* For example, nanoparticles in cosmetics or food with large scale production and high exposure rates.

## Second Generation

Active (evolving function nanostructures), for example, reactive nanostructured materials and sensors; targeted cancer therapies *From 2005 –*

*Potential risk:* For example, nanobiodevices in the human body; pesticides engineered to react to different conditions.

## Third Generation

Integrated nanosystems (systems of nanosystems), for example, artificial organs built from the nanoscale; evolutionary nanobiosystems *From 2010 –*

*Potential risk:* For example, modified viruses and bacteria; emerging behavior of large nanoscale systems.

## Fourth Generation

Heterogeneous molecular nanosystems, for example, nanoscale genetic therapies; molecules designed to self assemble *From 2015 to 2020*

*Potential risk:* For example, changes in biosystems; intrusive information systems (Renn and Roco 2006, p. 14; for a further discussion of the generations see Davis 2009).

The potential risks mentioned do not overlap completely with those to be discussed in this chapter. Risks of nanoparticles of course are mentioned explicitly in the first generation and are implicit in the second. While no mention is made of grey goo, it was thought to be a problem of self-assemble so would be a fourth generation issue. Cyborgs would fit into the third generation. Intrusive information systems, mentioned in the fourth generation, are related to the privacy question, but this is not only a future issue, it is here already, or so it will be argued. The nanodivide does not rate a mention.

## Nanoparticles

Most of the discussions of the risks of nanotechnology concern manufactured nanoparticles and if nanotechnology has any risks, many would consider these to be the only ones (for an overview see Seaton et al. 2010). In most cases, the issue is not that it is known that particular nanoparticles are health or environmental risks, but rather that not enough is known yet to understand their effects properly. Nanoscale particles, commonly said to those particles in the 1–100 nm range, have properties different from larger particles of the same material, due at least partly, to their greater surface area relative to size. According to Seaton et al. “Surface area is the metric driving the pro-inflammatory effects” and “In cells, high surface area doses appear to initiate inflammation through a number of pathways but oxidative stress-responsive gene transcription is one of the most important.” (Seaton et al. 2010, p. S123). However, their effects in the body are not yet well understood. This has led to calls for products containing these particles to be withdrawn for the market and a moratorium on further development until more research into their safety has been undertaken (FoE 2007). One concern is with products such as cosmetics and sunscreens that are applied to the skin. Some worry that the nanoparticles could pass through the skin and lodge in various parts of the body where they could cause harm.

Faunce et al., for example, suggest that the evidence so far is inconclusive whether or not titanium dioxide and zinc oxide, both used in sunscreens and known to be harmful to cells, can penetrate the skin (Faunce et al. 2008). Another concern is that threadlike particles that can be inhaled might have the same effects as asbestos and lead to serious lung disease (Bell no date). Nanoparticles in the air are of course nothing new and by and large cause no ill effects. The worry is that some manufactured particles of certain shapes might not be so benign (see Schmid et al. 2006, pp. 344 ff. for a detailed discussion). The use of nanotechnologies in food is also causing some concern and has been the focus of a recently published report in Great Britain. According to this report, potential application in the food industry include “creating foods with unaltered taste but lower fat or sugar levels, or improved packaging that keeps food fresher for longer or tells consumers if the food inside is spoiled” (House of Lords 2010, p. 5). It continues by saying that little is known of the potential dangers because of the paucity of research in the area.

## Grey Goo

---

Lest the mere mention of grey goo takes away all credibility from this chapter, let me say at the outset that it is being discussed, not because it seems to be a real possibility, but because it highlights some relevant aspects of risk. The potential problem of grey goo arises in the context of one kind of nanotechnology, molecular manufacturing, something that receives less attention in the literature now than it did a number of years ago. Molecular manufacturing is a “bottom up” approach where things are created or manufactured by the manipulation of atoms and molecules (see Drexler 2006; Treder and Phoenix 2007). Drexler (1996) envisaged tiny robots, or nanobots, that would have the ability to self-replicate indefinitely, and the grey goo problem would be the result if they were not suitably controlled. In this “bottom up” approach to nanotechnology, self-replicating robots could manufacture just about anything. The problem was seen to be that if the research continued, there would be a possibility of these nanobots actually being developed and therefore the possibility that they might escape from the environment in which they were developed for a particular task and consume everything around them without the possibility of containment. This “grey goo” risk seems now to be minute because of both theoretical and practical problems with certain approaches to molecular manufacturing (see Smalley 2001; Drexler 2001; Phoenix and Drexler 2004).

The issue here is that while the risk appears to be minute and certainly not a short or medium term one, the results could be catastrophic. Whether such risks should be considered at all is a moot point. If they are only possible in the very distant future, the risks of biological and nuclear weapons used in war or terrorism, or environmental degradation, and extreme climate change seem likely to decimate us before grey goo gets its chance.

Richard Posner argues, however, that even if the risk is tiny, given that the result would be catastrophic, the risk should be taken seriously. He expresses skepticism about the scientists’ claims that it will almost certainly never happen “given the record of scientists’ ‘never’ predictions” (Posner 2004, p. 36).

## Privacy

---

New monitoring and surveillance technologies are not commonly classified as risks, but they do pose threats to personal privacy so it is not unreasonable to classify them in this way.

They are not risks in the same way as the potential toxicity of some nanoparticles are risks to health or the environment. There is, however, the potential for harm through loss of autonomy and increased vulnerability to control. Moreover, just as manufactured nanoparticles are already in use and some evidence exists, as seen earlier, of dangers, so technologies such as certain computer technologies are already threatening privacy. These new monitoring and surveillance technologies therefore do pose risks. The problem is that the more that is known about a person, the greater the ability of those with the information to harm that person. This is the reason for strict controls on certain kinds of personal information, for example medical records. It is not so much that it necessarily matters what others know about one. The central issue is how the information that they do have is used.

That developments in nanotechnology, particularly in nanoelectronics, will enhance monitoring and surveillance techniques and capabilities is almost certain. Faster processing of increasing amounts of data on smaller devices and evermore powerful and sensitive sensing devices will ensure this. These technologies are not being developed, for the most part, with the explicit aim of reducing personal. The drivers are increased productivity and efficiency (e.g., RFIDs in warehouses, Kelly and Erickson 2005), and particularly security against potential terrorist attacks.

## Cyborgs

---

Technological and scientific developments enabling human enhancement employ many techniques, perhaps the most prominent being genetic engineering and the use of drugs. In one aspect of enhancement, however, nanotechnology does have an important role, and that is the development of cyborgs; the merging of humans and machines. Computer technology is of course vital here but developments in nanoelectronics to further miniaturize computing and sensing devices and the creation of new materials will be an essential element in the continued research and development of cyborgs (see Roco and Bainbridge 2002, for an early discussion).

The merging of machines with humans poses risks in a number of ways. It might threaten human integrity and it might have adverse effects on individuals or communities or both. This case is rather different from the previous ones. In the first two the potential harm is physical with no doubt that it is harm. The third is a little more contentious but a reduction in privacy and autonomy usually must be justified by an increase in other goods, for example security. So, other things being equal, there is a perception that their loss is regrettable. It is less obvious, however, that loss of human integrity is a harm and even if it is, that combining technology with humans is a cause of the loss. The lack of clarity is at least partly because human integrity is not easy to define. It depends on beliefs about what it is to be human. It is not normally thought that someone who wears spectacles, contact lenses, a hearing aid or has a pacemaker is less human because of that reliance on technology. The worries occur with technological enhancements rather than technologies for therapy. It is enhancements that threaten human integrity. This therapy-enhancement distinction, however, does not stand up well to close scrutiny. The claim then that the technologies in question pose a risk to human integrity is contentious and depends both on what it is to be human and what constitutes enhancement. But nevertheless, it is considered by many to be a risk. The risk of other harms to individuals and communities is also contentious but more comprehensible. Just two will be mentioned here. Jürgen Habermas (2003)

raises the concern that if certain sorts of enhancements are available, parents will feel obliged to have their children enhanced just as today they feel obliged to give their children the best education that they can afford. The problem with enhancements is that they may not be reversible and the children may grow up with enhanced abilities that they do not want and therefore have less autonomy to do what they really want to do. Another issue is that if a range of enhancements is possible and different people are enhanced in different way, then communication may become difficult between these different people and social cohesion will suffer (see Savulescu and Bostrom 2009; Lin and Allhoff 2008 for further discussion).

## Nanodivide

---

A final worry is that nanotechnology will help increase the divide between the developed world and the developing world. The term “digital divide” is used to name the gulf between those with and those without adequate information and communication technologies. Likewise, “nanodivide” names what some see as the potential for nanotechnology to further disadvantage poorer peoples relative to the richer. Most of the technologies will be for those already well-off. Those are the people who can afford to buy new technology and so products will be developed for them. These harms to the developing world are relative; the poorer parts will become relatively poorer simply because the richer parts will be richer in absolute terms. Another aspect of the nanodivide concern is that products will be developed and produced in rich countries that replace natural products now produced in poorer countries. This concern was raised by the ETC group (2004) in relation to synthetic rubber production that would replace natural rubber. The creation of a nanodivide is not seen as a risk by all. Some argue, for example Peterson and Heller (2007), that nanotechnology will in fact provide extensive benefits to developing countries, but others, for example Joachim Schummer (2007), show cause to doubt this.

Before looking at whether scientists or the scientific community more generally is morally responsible for harms resulting from these risks, we consider the scientific enterprise itself. This will provide a background for the discussion.

## Models of Science

---

To a large extent the model that is held of scientific research determines what constitute the moral responsibilities of scientists. Two such models will be considered, the linear and the social models. These can be interpreted as either descriptive, that is as science is actually done, or as normative, that is, as it ought to be done. Here it will be suggested that as a description of how science is actually done, the linear account is not accurate and that the social model is the better one from the moral standpoint. This will become clearer when the role of values in science is discussed.

### The Linear Model

---

This model, it is often said, dates back to Francis Bacon and sees a progression from pure science through applied science and technology, to products. It maintains that scientists are the

best judges of what research should be pursued and that society benefits most when the scientists are given free reign to follow the search for knowledge wherever it may lead. A strong advocate was Michael Polanyi, who wrote in 1962:

- ▶ During the last 20–30 years, there have been many suggestions and pressures towards guiding the progress of scientific inquiry in the direction of public welfare.

...

Any attempt at guiding scientific research towards a purpose other than its own is an attempt to deflect it from the advancement of science. . . . You can kill or mutilate the advance of science, you cannot shape it (Polanyi 1962).

Another enthusiastic supporter was Vannevar Bush who had a large influence on science policy in the USA. He wrote:

- ▶ Scientific progress on a broad front results from the free play of free intellects, working on subjects of their own choice, in the manner dictated by their curiosity for exploration of the unknown. Freedom of inquiry must be preserved under any plan for Government support of science . . . (Bush 1945).

This view still has currency, at least amongst many scientists, and seems to be that held by the Australian scientist Sir Gustav Nossal who, speaking of scientific research, writes:

- ▶ To free the human spirit from ignorance and superstition must be good, no boundaries should be placed around such a search (Nossal 2007, p. 6).

Much research, however, eventually leads to technologies useful to, or possibly detrimental to, individuals and society, therefore:

- ▶ This is why the distinction between science (which seeks to know) and technology (which seeks to apply knowledge) is so crucially important. Of course technology must be subjected to societal and democratic norms (Nossal 2007, p. 7).

He seems to imply here that scientific research should not be subject to such norms, that is, that scientific research is value-free with respect to moral values.

In summary, this model asserts that science progresses best when scientists are given free reign to follow their research interests and that this also leads to the best results for society.

## The Social Model

---

The linear model is held by many, especially scientists, but is not unchallenged. It can be argued, as already noted, that not only is it not the way that science works, that is, it is factually wrong, but it is not even the best way for science to work; it is not something at which we should aim. Albert Einstein reflects this view in a talk to students:

- ▶ It is not enough that you should understand about applied science in order that your work may increase man's blessings. Concern for man himself must always be our goal, concern for the great unsolved problems of the distribution of goods and the division of labor, that the creations of your mind may be a blessing, and not a curse, to mankind. *Never forget this in the midst of your diagrams and equations* (Einstein 1931).

The social model differs from the linear in a number of ways. Most importantly, on the social modal science is not value-free and this will be discussed in the following section. Here, we will focus on two aspects of this model: first, the involvement of society in decisions about what research should be undertaken, that is, scientists should not have completely free reign; and second, that there is no sharp distinction between pure science on the one hand and applied science and technology on the other.

First, should society have a role in setting the research agenda? The European Commission has recently published two documents that advocate broader society involvement in decisions on what research should be undertaken. In their “Code of Conduct for Responsible Nanosciences and Nanotechnologies Research,” one of the clauses states:

- ▶ 4.1.13 Member States, N&N research funding bodies and organizations should encourage fields of N&N research with the broadest possible positive impact. A priority should be given to research aiming to protect the public and the environment, consumers or workers . . . (Commission of European Communities 2008).

This clause clearly advocates a research agenda that gives the greatest benefits to the greatest number of people. More importantly for our purposes, it advocates a research agenda that is not determined just by scientists’ interests or by profits. It introduces a social element. Similarly, in the *Global Governance of Science* report it is argued that:

- ▶ In a world of competing goods and limited resources – in which sciences not the only good and all research programs are not equally able to be funded – the governance of means must be complemented by a governance of ends (Global Governance of Science 2009, p. 41).

In other words, the curiosity of scientists is not enough to set the research agenda; the results, or ends, of the research must be taken into account.

In that report, John Dewey is referred to approvingly as someone who advocated this broader view of science. While it is true that he does maintain that scientists have a “supreme intellectual obligation” to society, and says, for example:

- ▶ The wounds made by applications of science can be healed only by a further extension of applications of knowledge and intelligence; like the purpose of all modern healing the application must be preventative as well as curative. This is the supreme obligation of intellectual activity at the present time. The moral consequences of science in life impose a corresponding responsibility (Dewey 1934, p. 98).

His position, however, is a little more subtle than stated in that report. One of his main arguments is that the “supreme intellectual obligation” is to develop the ability to think scientifically, critically, and rigorously, which in turn will benefit society.

If research is aimed at improving life, whether it be better health, cleaner energy, greater profits, or whatever, serious questions must be asked concerning the kind of research that should be done and who should make the decisions. With respect to publicly funded research, the decisions are commonly made by governments, at least broadly by setting research priorities. In Australia for example, the priority funding areas in 2011 are:

- An environmentally sustainable Australia
- Promoting and maintaining good health

- Frontier technologies for building and transforming Australian industries
- Safeguarding Australia (ARC 2009).

Philip Kitcher, however, argues that governments are not the best bodies to decide what research should be done, but neither does he believe that, in a democracy, all decisions should be left to the scientists either. His suggestion is this:

- ▶ . . . we need an institution that would offer a serious map of what scientific research has achieved and what possibilities it opens up for us. There's no reason why a group of senior scientists, perhaps scientists nearing the end of distinguished research careers, could not produce, in joint deliberation, a relatively accurate overall picture. They could then present that picture to representatives of different human constituencies in different human societies, representatives who would thus come to understand the lines along which future inquiry might be conducted. This *Scientific Forum*, . . . could then be set the task of trying to devise an appropriate research agenda that would take seriously the tutored preferences of all the represented groups. . . . perhaps under the aegis of UNICEF (or UNESCO) or some similar group, . . . (Kitcher 2007, pp. 183–184).

Kitcher's argument is a general one that applies to all sciences. It can be questioned whether his suggestion is practical. Often it may only be the scientists who are involved in the research who understand enough to really know what the future direction of the research should be. This is probably true with respect to the details of the research agenda but at a higher, "big picture" level his suggestion has plausibility. If Kitcher and the European Commission are correct about the desirability of research agendas being set with social goals in mind and not merely as a result of scientists following their own interests, or of profits, then of course there must be consideration of those social goals. This is one place that moral values enter into scientific research. We turn now to these values as a way of highlighting some important issues relating to responsibilities for risks.

The second difference between the linear and social models mentioned earlier is that the former but not the latter wants to maintain a strict division between pure science on the one hand and technology on the other. While there are obviously differences in degree, as described, for example, by Mario Bunge (1988), any sharp difference is difficult to maintain. Kitcher (2001) supports this view and argues for "the myth of the pure" (my emphasis), that is, that it is a myth that some science is so pure that it is not contaminated by applied science or technology. Advocates of technoscience also see no sharp distinction (Hottois 2005). If it is true that science and technology are closely related there are at least two consequences for the argument in this chapter. First, ethical and social values permeate science to a greater degree than is often recognized and second, it has implications for the responsibility of scientists.

Just one argument for the relationship will be given here that has relevance for scientific responsibility (for others see Kitcher 2001). We will assume that the results of most, if not all, successful scientific research will be published. Once it is published it is in the public domain and can be used in future research by anyone. We can assume too that most successful research that leads to discoveries will have uses apart from merely increasing human knowledge and reducing ignorance. If this is so, it is almost certain that those discoveries will be used in the development of new technologies and that these technologies will be used, hopefully for good but possibly also for harm. This link between science and technology is not a logically necessary one and there is nothing deterministic in any strong sense, about the science leading to the technology. It is not logically inevitable, but given what we know about the way that people

typically behave, particularly in a capitalist society, there is a kind of factual inevitability about the link. This factual inevitability is underpinned by a normative inevitability that resembles what Bruce Bimber calls normative technological determinism (Bimber 1994). It is not determinism at all but has that appearance because the technological development is driven largely by the unquestioned values of efficiency and productivity so always moves in the direction that those two values direct. In the science-technology link, publishing is driven by various values: desire to increase knowledge or to improve the world in some other way; continued or future employment; prestige of one's institution or oneself. Once the knowledge is available to all, if it is at all useful the values of productivity and efficiency will almost ensure that technology will be developed and used.

## Values in Research

---

In order to understand better the place of moral responsibility in each model, it is useful to look at the role of ethical issues in each a little more closely. Each of these models has some different ethical implications. In order to explore these, it will be useful to outline the place of values in scientific research.

It is sometimes claimed by defenders of the linear model that research, especially pure research, is value neutral with respect to moral values, and that values enter only at the level of technological development, something that we have already questioned. What is meant is that while the results of the research can be used for good or ill and so are value-laden, those values do not permeate the research itself. That research merely generates new knowledge. It is in the use of that knowledge that values reside.

This view, it seems, is not accepted by Nossal:

- ▶ I have always believed that scientific research is richly value-laden, that the search for new knowledge gives expression to a deep human yearning, that seeking to know more about the natural world is an unalloyed good (Nossal 2007, p. 6).

The values acknowledged here are satisfying human yearning for knowledge and gaining knowledge itself. While research is value-laden in this sense, it is not what is commonly meant when it is claimed is that scientific research is value-laden with moral values.

Obviously, broader social values come into play at the stage of technological development, particularly pertaining to the products that should be developed, to their design, and to their use.

The social model incorporates many of these latter values into the scientific enterprise. At the very least, research should be directed to human needs more generally and not limited to increasing knowledge. It is at this stage that the difference of ethical import of the models lies. The tentative suggestion is that at the research stage on the linear model the main value is increasing human knowledge; on the social model, other social values are equally and perhaps even more important.

But this is a preliminary conclusion only and slightly misleading at that. Nossal's values, it must be noted, are not strictly part of scientific research itself but are values that underlie, or lead to, at least some kinds of research. The same could be true of the social values mentioned. Moral and social values are important in science overall, it is sometimes argued, but they are not part of science proper. If this is so then the linear and social models are not incompatible because they are focusing on different things; the linear model on the research itself and the social on the scientific enterprise more generally.

In order to spell out in more detail the relationship between ethical values and science, we will consider a number of areas where they potentially meet, called here interfaces.

## The Science/Ethics Interface

---

Ethical concerns arise at various stages of the scientific enterprise and here four interfaces between the two will be considered.

### Interface 1

---

Ethics in science is often limited to what ethics committees oversee and this is largely the way that research is undertaken: will the research harm human or animal subjects, is it safe for the researchers, are intellectual property rights being respected, and so on. However, this is only one interface between ethics and science, and perhaps the only one that defenders of the linear model will accept.

### Interface 2

---

A second, and contested, interface is the relationship between values and decisions to accept or reject hypotheses. Richard Rudner (1953) criticizes the view that no values are part of the research itself. His central argument is that a core part of scientific research is to accept or reject hypotheses and that this decision making cannot be divorced from the scientist's values. Values must come into play because risks of harmful consequences of certain theories. The greater the potential risk the greater must be certainty that the hypothesis is true so the scientist must make a value judgment regarding the degree of certainty. Is the certainty great enough to warrant accepting the hypothesis? So scientific research *qua* scientific research is not value neutral.

A potential problem with Rudner's argument is that it is not clear that the value judgments enter at the level of hypothesis acceptance or rejection. The commitment to certain standards of rigor for acceptance or rejection could be a value stance taken prior to undertaking the research, an argument advanced by Hugh Lacey (1999). According to Lacey, deciding what level of certainty is appropriate in a particular instance is a questioning of the generally accepted standard rather than a decision based on a value judgment as part of the research itself. What Rudner has shown is another of the "important aspects of the 'touch' (or ... 'constant rubbing') of science and values" (Lacey 1999, p. 73). This is close to what in this chapter is called the *interface* between science and ethics.

One problem with Lacey's approach is that the actual value free domain of science is extremely narrow. Another and more important one is that it does not avoid Rudner's objection to decisions being free of value judgments. The generally accepted standards that are being questioned might be outside the research proper, but the questioning itself is part of the role or scientist *qua* scientist. If the scientist is worried about the risks of being wrong, his or her rejection of the standards in this case might be well founded, and clearly based on a value judgment regarding acceptable risk.

A similar point is made Heather Douglas (2009) in her much more nuanced account of values in science. Douglas considers three categories of values: ethical, social, and cognitive. All have a role to play, albeit, most of the time an indirect rather than a direct one. While ethical values should not directly influence which hypotheses should be accepted, for example, a vegetarian scientist should not let his or her vegetarianism influence the acceptance or rejection of an hypothesis about the benefits of meat eating, ethical values can be important indirectly, for example, in cases where the scientist must make a judgment regarding certainty given high risk, as Rudner argues.

### Interface 3

---

It was noted earlier that Nossal argued for the value-ladenness of scientific research. Research has built-in values; satisfaction of a human need and the goodness of expanding knowledge. This is another interface and one where the two models diverge from an ethical standpoint. On the social model, Nossal's two values are not enough and could be overridden by other social values. Research should be for the good of society more generally and the agenda for research should be set with priorities, some projects being much more important and urgent than others. For example, in the situation where the burning of fossil fuels is increasing levels of carbon dioxide in the atmosphere to a worrying degree, nanotechnology research perhaps should have clean energy as a much higher priority than, say, research into better sun screens or drug delivery devices (Weckert 2010). On this model, the priorities should be set by the government or some other bodies representing society generally, as mentioned in the earlier discussion of the social model.

### Interface 4

---

The fourth interface concerns responsibility. Which researchers are responsible for benefits or harms of research will depend, at least partly, on which model is held. There is little argument that use of technology carries responsibilities with it. If I use some technology in a way that harms others I can be held morally responsible. Responsibility for technological development is more contentious. If I develop some product that is used in harmful ways my responsibility for the harm is less clear. I can always fall back on the claim that the technology is not the problem; rather the problem is in the use. Finally, the responsibilities of scientists who undertook the research that enabled the technology to be developed and used for harm are more contentious again. In fact, it is commonly denied that they carry any responsibility at all for such harm. Their responsibilities are limited to issues of doing the research well. This account of scientific responsibility is closely tied to the linear model where there is a clear distinction between science and technology, something dismissed earlier. The distinction is seen as the boundary for moral responsibility and legitimate societal or governmental control or interference.

On the social model, responsibility goes all along the chain, back to even the pure science, something that does seem to be implied in statements above. Values external to the science itself come into play right at the beginning. Those further removed from the consequences will have diminished responsibility but they cannot avoid it completely. This of course leads to what can be seen as a completely untenable position and perhaps even a *reductio ad absurdum* of the

proposed view; given the dual use dilemma that most research can be used for good or harm (Selgelid 2009), scientists are morally blameworthy then, even when their research is undoubtedly intended to promote nothing but good. They would be blameworthy regardless of what they did, an absurd position. A closer look at moral responsibility shows that the situation is not quite so dire although things may be a little less comfortable than on the linear model.

The discussion of this section has focused on a number of points where moral values become important in science and highlighted differences between the linear and social models with respect to these values. The differences largely are a function of what is taken to be the scope of science. The linear model restricts it just to the research activity itself whereas on the social model the research in its societal context is the focus. Values relating to what research should be undertaken and to the effects of the research on society therefore become important and essential parts of science itself. This difference clearly has implications for scientific responsibility, as seen in the previous argument. More must be said now about responsibility.

## Moral Responsibility

---

The word “responsibility” has two distinct senses that are important for our purposes. First, there is the *causal* sense: if it is said that lightning was responsible for a fire, what is meant is that lightning caused the fire. It is ambiguous, however, to say that someone was responsible for fire. This might just mean that that person caused it or it could mean that he or she is praiseworthy or blameworthy for lighting it. This second sense is *moral* responsibility. Someone is morally responsible for an action if he or she played a causal role in the action and if it is appropriate to attribute praise or blame for causing it. A person who is responsible in the causal sense is not always responsible on the moral sense.

In order to be held morally responsible for some harm, a person must at least (1) have caused it or knowingly allowed it to happen when it could have been prevented and (2) must have intended to cause it, or allowed it through negligence or carelessness. Condition (1) would frequently be satisfied on the grounds that there is a causal link between some research and some harm. Without the research, the harm would not have been possible. Condition (2) however, hopefully is seldom satisfied. Scientists rarely *intend* to cause harm. The second part of the condition though gives some pause for thought. A scientist cannot avoid all moral responsibility simply because no harm was intended. If no, or insufficient, thought is given to possible harmful consequences, some moral responsibility would be present.

It might be objected, reasonably, that this picture is overly simplistic because research now is not usually conducted by just one person. Normally it will be undertaken by a team, possible a very large one. In such cases, the argument goes, nobody can be held morally responsible. This is the problem of collective responsibility, a problem that arises when groups cooperate to achieve some end, whether it be scientific research, conducting a war, or developing computer software (Miller 2008).

Because many people are involved in typical scientific research projects, when something goes wrong, it is not always easy to say who is morally responsible, and who, if anyone, ought to be held accountable and liable for any damages. One solution is just to say that the group, or organization, is responsible. In everyday talk we do this frequently, a good example being British Petroleum’s (BP) responsibility for oil spill in the Gulf of Mexico in 2010. Two problems are worth pointing out here. One is a degree of unfairness. Not everyone in a group is equally

responsible or even responsible at all. Not all BP employees played a role on the explosion and spill. The second problem is that moral responsibility is something that can only be attributed to autonomous human beings, not organizations.

While it is probably true that in most cases where there is collective responsibility no individual will bear the entire blame, it does not follow that moral responsibility cannot be attributed to many individuals. Many can therefore be held accountable to varying degrees. Not everyone in a research or development team will necessarily bear the same level of responsibility for faulty products or other harms resulting from the research and development, and some perhaps will have no responsibility at all. A careless researcher or developer can bear responsibility as can someone who is negligent in the testing of products or harmful uses of research outcomes. In most cases, ultimate responsibility must be borne by the leaders of projects whose role is general oversight the whole process. Responsibility cannot be avoided simply because a large team is involved.

A further aspect of collective responsibility must be noted. While not all individuals are at fault and those that are, not all equally so, there is a sense in which all, with only several exceptions, can plausibly be held accountable. If I do nothing to try to change the situation of my team's or company's carelessness, I am helping to perpetuate a climate in which faults, mistakes or accidents are more likely to occur. So I can and ought to be held accountable to some extent, even though I did not cause the events, or intend them to happen. In Larry May's terms, I can be morally tainted even if I cannot be blamed for the event itself (May 1991). The exceptions are where I have protested or attempted to change the situation, or am not in a position where I could do anything, for example, if I could not reasonably be expected to know of the situation. Given this, all, or most, members of a group, team or company can be held collectively responsible for research.

## Risk and Moral Responsibility in Nanotechnology

---

In the first section of this chapter, various risks or potential risks associated with nanotechnology were outlined. In the second, two models of science were considered and each of these had different implications for scientific responsibility. Next, a little more was said about moral responsibility itself. In this section we will return to the five areas of risk discussed in the first section and in each case examine the responsibility for the management of the risks and for harms caused. Various levels can be distinguished for moral responsibility in these cases: scientific researchers, developers of technologies, users of the technologies, and regulators. Our primary interest is in the moral responsibilities of scientists and the scientific enterprise generally.

### Risk with Nanoparticles

---

As we saw earlier, the main focus of discussions of risks in nanotechnology has been on those associated with the potential toxicity of various manufactured nanoparticles. Given that the particles are at the core of the research, a strong case can be made that some moral responsibility attaches to the researchers if those particles cause harm to health or the environment. If praise for the benefits of this research is appropriate, then blame for the harms is too. But of

course it is not so simple. The research, it is fair to assume, is aimed at the benefits whereas the harms are unfortunate and unintended consequences. If the benefits derived from the particles cannot be realized without the potential for harm, then one primary responsibility of the scientists is to make clear what the potential risks are and another is to undertake further research to see if the risks are real and, if they are, to try to find ways to mitigate them (see Maynard 2006) on addressing risks. While the scientists are attributed more responsibility in the social model, even in the linear one they do not escape completely. An important scientific responsibility in this model is to increase knowledge, and increasing knowledge of the toxicity of nanoparticles and their movement and accumulation in human bodies and the environment is a vital part of this knowledge.

More direct moral responsibility can be attributed to developers and manufacturers of products containing potentially toxic nanoparticles, providing of course they have been given sufficient information from the researchers about potential toxicity. Take sunscreens for example mentioned earlier. Some contain nanoparticles and debate continues over their safety. Do the particles penetrate the skin and if they do, what harm do they cause? If harm is indeed caused then the products should be taken off the market (Faunce et al. 2008), and if they are not, those producing them are morally responsible for the harms. But again, just as in the case with the researchers, more must be said. If the producers are not warned by the researchers then they can hardly be held responsible providing that they have taken measures to find out. If there are no products that are anywhere near as effective in preventing skin cancer and other sun related problems, then the risk may be worth taking and they bear little, if any, responsibility. Another issue is labeling of the products. If the labels state clearly that the product contains nanoparticles with a warning of potential risks, then much of the responsibility is shifted onto the user. This suggests that some of the responsibility also lies with regulators. They have a responsibility to warn the public of dangers. This of course raises another issue, that of difficulties in regulating nanoparticles or products containing them.

## Risk with Grey Goo

This case is different from the previous one in three respects. First it is extremely unlikely, second it is not a short or medium term concern and third, the consequences would be catastrophic. If it did occur and the consequences were as predicted by some, no question of moral responsibility would arise simply because nobody would be left alive. The responsibility question then is whether scientists have a responsibility not to undertake research into self-replicating robots that could lead to grey goo, or at least whether there is a responsibility to undertake research into controlling the self-replication process simultaneously with the research into developing it. The second of these is the more plausible given that the first would deny the world of any benefits as well. On the social model this is clearly true, but even on the linear, as in the previous case, there is a responsibility to increase knowledge of the dangers and how to mitigate them.

This leads to a final consideration is the “If we don’t someone else will” argument. One version of this is something like “this activity might be harmful or otherwise morally dubious, but it will be done by someone so we might as well do it. It is unfair to hold us responsibility for the harms because they would occur whether or not it did it.” This can easily be used as an excuse but other versions have more plausibility. Suppose that we would do it more carefully

and safely than others so that any harmful consequences are reduced? An even stronger argument can be made. Perhaps, there is a moral responsibility for us to undertake the activity, in this instance the research, *in case* others do so because if they succeed we have a way of controlling the nanobots if something does go wrong. Undertaking the research may in fact be the best way of managing the risks even though it might be better if the research were done by nobody at all. This “If we don’t someone else will” argument is more complex than the account given here and is most plausible on consequentialist grounds. Much more needs to be said, but for the purposes here, there are to be instances where it is justified and this may be one of them (Glover 1975; May 1991).

## Privacy

---

Personal privacy, and therefore autonomy, is being threatened by new technologies, including nanotechnology. The risks are real, not merely potential although they have the potential to become much worse. Results of scientific research led to the development of various technologies that enhanced capabilities for monitoring and surveillance and therefore to the threats to personal privacy and the harms that can follow from that. The scientists therefore have a direct causal link with the harms but the claim that they also have moral responsibility is more tenuous and on the linear model they probably have none at all. Monitoring and surveillance have many legitimate uses and the more effectively and efficiently they are carried out, the better in many cases. This is a standard dual-use case where research is undertaken to produce some good, in this case, legitimate surveillance, but then is used for ill, illegitimate surveillance, as well. Another factor is that illegitimate surveillance in itself does not necessarily cause harm, rather it is the use that is made of the information gained that is often used for harm. Moral responsibility here rests primarily with users of the technology. It is less clear here than in the previous examples how more scientific research could improve the situation.

## Cyborgs

---

As discussed earlier, the risk here is primarily to human integrity, if a risk exists at all. If human integrity suffers then those undertaking the research do bear moral responsibility. Because the purpose of the research is to enhance human performance in some way by the integration of human and machine any harm to integrity cannot be claimed to be an unintended consequence or a result of dual-use. Loss of integrity in this instance is constituted in combining human and machine. The best way to manage the risk therefore is not to do the research in the first place. This way of looking at it, however, hides two important considerations. The first is that useful medical devices result from this kind of research and second, a distinction must be drawn between research into the technology itself and that into linking the technology and humans. The first point draws on the therapy-enhancement distinction. If this is a viable distinction, then research into therapy is justifiable, and if the same research is used for enhancement, then this is a dual-use issue and the researchers bear no responsibility. The distinction, however, is the subject of much dispute so reliance on it does not solve many issues. The second point concerns different kinds of research required for human enhancement. Researchers working on computing or other necessary technology bear little if any responsibility.

Even if their work led to the development of computers onto which the contents of human brains were downloaded, their responsibility is limited to the development of the computer, not the uses to which it is put. However, those engaged in research into how best to link humans and computers for the purposes of enhancing capabilities do bear responsibility. The mitigating consideration here is that it is not clear that loss of human integrity occurs or that any harm at all results from this enhancement.

## Nanodivide

---

It is difficult to attribute moral responsibility for any nanodivide to those involved in research into and development of technologies and products creating that divide. It is primarily a political problem. The research is done and technologies developed and if the distribution of the resulting benefits is unfair, that is hardly the fault of the researchers and developers. The situation, however, is not so clear where the research benefits the well-off only or benefits them at the expense of the less affluent. In medical research 90% of funding goes to diseases of 10% of the world's population and a similar situation occurs with private funding of agricultural research. It could be argued that researchers do have a responsibility to do more to help the less developed parts of the world. Furthermore, some research will probably be detrimental to some poorer areas, for example into synthetic rubber. While in these cases some responsibility may lie with the scientists, governments and other research funding bodies bear a greater responsibility.

## Conclusion

---

As with all technologies, nanotechnologies come with risks as well as benefits. Scientists clearly have a responsibility to help alleviate some of the risks, particularly with respect to nanoparticles. They have the twin responsibilities of undertaking the research and insisting that funding is available to do it. If molecular manufacturing becomes a reality they will have similar responsibilities to ensure that appropriate research is done to minimize its risks. Scientific responsibility is less in the other risks areas discussed in this chapter. It is not so much extra research that will lessen risks to privacy, but rather restrictions on the uses of developed technologies and products. Whether or not cyborgs pose risks depends largely on views of human nature and nanoscientists, as scientists, cannot do much about that. This situation is similar with respect to the nanodivide, although in this case it can be argued that more research should be undertaken in areas that are likely to decrease the divide. If we want the benefits of nanotechnology, we must cope with the risks, and the scientists whose research generates those risks have an important role to play in the management of at least some of them.

## Further Research

---

With the continued development of technologies such as nanotechnology, synthetic biology, and genetic engineering, with their potential for altering not only the environment but also

human beings, further research is required to understand better the consequences of these developments. The research into social and ethical issues must be interdisciplinary and based on sound science and a knowledge of actual technological developments and realistic assessment of potential developments. A strong focus on what is required from science and technology is necessary with respect to improving life for all if harms are to be minimized and benefits maximized. This research has begun but more is required, particularly in relation to understanding responsibility for harms and the role of technology in the good life.

## Acknowledgment

I would like to thank the Ethics Centre of South Australia and the Ian Wark Research Institute (Particle and Material Interfaces) at the University of South Australia, for giving me a Research Fellowship in Ethics in 2008. The discussions there were valuable in the formation and development of the ideas in this paper.

## References

- ARC (2009) Australian Research Council. [www.gov.au](http://www.gov.au). Accessed 20 May 2011
- Bell TE (no date) Understanding risk assessment of nanotechnology. Article funded by National Nanotechnology Coordination Office. [http://www.nano.gov/Understanding\\_Risk\\_Assessment.pdf](http://www.nano.gov/Understanding_Risk_Assessment.pdf). Accessed 14 March 2011
- Bimber B (1994) Three faces of technological determinism. In: Smith MR, Marx L (eds) Does technology drive history: the dilemma of technological determinism. MIT Press, Cambridge, MA, pp 79–100
- Bunge M (1988/2001) The nature of applied science and technology. In: Maher M (ed) Scientific realism: selected essays of Mario Bunge. Prometheus Books, Amherst (Chap. 24)
- Bush V (1945) Science the endless frontier. United States Government Printing Office, Washington, DC
- Commission of European Communities (2008) Commission recommendation on a code of conduct for responsible nanosciences and nanotechnologies research, Brussels, 7 Feb 2008, C(2008)424 final
- Davis JC (2009) Oversight of next generation nanotechnology. PEN 18 Apr Woodrow Wilson International Center for Scholars, project on emerging nanotechnologies. <http://www.nanotechproject.org/publications/archive/pen18/>. Accessed 20 May 2011
- Dewey J (1934) The supreme intellectual obligation. *Science* 79:240–243
- Douglas H (2009) Is science value free?: science, policy, and the value-free ideal. University of Pittsburgh Press, Pittsburgh
- Drexler KE (1996) Engines of creation: the coming era of nanotechnology. Fourth Estate, London
- Drexler KE (2006) Nanotechnology: from Feynman to funding. In: Hunt G, Mehta M (eds) Nanotechnology: risk ethics and law. Earthscan, London, pp 25–34
- Drexler KE (2001) Machine-phase nanotechnology. *Sci Am* 285:66–67
- Einstein A (1931) Address before student body, California Institute of Technology, 16 Feb 1931. Quoted in McGinn, E. Ethical responsibilities of nanotechnology researchers: a short guide (2010) Nanoethics: ethics for technologies that converge at the nanoscale 4:1–12
- ETC Group (2004) Down on the farm: the impact of nano-scale technologies on food and agriculture. [www.etcgroup.org/upload/publication/80/02/etc\\_dotfarm2004.pdf](http://www.etcgroup.org/upload/publication/80/02/etc_dotfarm2004.pdf). Accessed 20 May 2011
- Faunce T, Murray K, Hitoshi N, Bowman D (2008) Sunscreen safety: the precautionary principle, the Australian therapeutic goods administration and nanoparticles in sunscreens. *NanoEthics* 2:231–240
- FoE (2007) International Union of Food, Farm & Hotel Workers considers global nano-moratorium. Friends of the earth. <http://nano.foe.org.au/node/185>. Accessed 20 May 2011
- Global Governance of Science (2009) Report of the expert group on Global Governance of Science to the Science, Economy and Society Directorate, Directorate-General for Research, European Commission, Ozoliņa Ž, Chairwoman, Mitcham C and Stilgoe J, Rapporteurs

- Glover J (1975) Part 1 of Glover J and Scott-Taggart M it makes no difference whether or not I do it. *Proc Aristo Soc Suppl* 49:171–209
- Habermas J (2003) The future of human nature. Polity Press, Cambridge
- Hottois G (2005) Technoscience (trans: Lynch JA) In: Mitcham C (ed) Encyclopedia of science and technology ethics. Thomson Gale, Detroit, pp 1914–1916
- House of Lords (2010) Nanotechnologies and food, vol 1. Great Britain: Parliament: House of lords: science and technology committee report
- Kelly EP, Erickson GS (2005) RFID tags: commercial applications v. privacy rights. *Ind Manage Data Syst* 105:703–713
- Kitcher P (2001) Science, truth and democracy. Oxford University Press, Oxford
- Kitcher P (2007) Scientific research – who should govern? *Nanoethics* 1:177–184
- Lacey H (1999) Is science value free: values and scientific understanding. Routledge, London
- Lin P, Allhoff F (2008) Untangling the debate: the ethics of human enhancement. *Nanoethics* 2:251–264
- May L (1991) Metaphysical guilt and moral taint. In: May L, Hoffman S (eds) Collective responsibility: five decades of debate in theoretical and applied ethics. Rowman and Littlefield, Savage, pp 239–254
- Maynard AK (2006) Nanotechnology: a research strategy for addressing risks. Woodrow Wilson International Center for Scholars, PEN 3
- Miller S (2008) Social action: a teleological account. Cambridge University Press, Cambridge
- Nossal Sir G (2007) In: Mills J (ed) Introduction, ethically challenged: big questions for science, The Alfred Deakin debate. The Miegunyah Press, Melbourne
- Peterson C, Heller J (2007) Nanotech's promise: overcoming humanities most pressing problems. In: Allhoff F, Lin P, Moor J, Weckert J (eds) Nanotechnology: the ethical and social implications of nanotechnology. Wiley, Hoboken, pp 57–70
- Phoenix C, Drexler E (2004) Safe exponential manufacturing. *Nanotechnology* 15:869–872
- Polanyi M (1962) The republic of science: its political and economic theory. *Minerva* 1:54–74
- Posner RA (2004) Catastrophe: risk and response. Oxford University Press, Oxford
- Renn O, Roco M (2006) Nanotechnology risk governance. White Paper No. 2, International Risk Governance Council, Geneva
- Roco MC, Bainbridge W (2002) Converging technologies for improving human performance: nanotechnology, biotechnology, information technology and cognitive science. A National Science Foundation/Department of Commerce sponsored report, Arlington. [www.wtec.org/ConvergingTechnologies/Report/NBIC\\_frontmatter.pdf](http://www.wtec.org/ConvergingTechnologies/Report/NBIC_frontmatter.pdf). Accessed 20 May 2011
- Rudner R (1953) The scientist qua scientist makes value judgments. *Philos Sci* 20:1–6
- Savulescu J, Bostrom N (eds) (2009) Human enhancement. Oxford Press, Oxford
- Schmid E, Ernst H, Grunwald A, Grüwald W, Hofmann H, Krug H, Janich P, Mayor M, Rathgeber W, Simon U, Vogel V, Wyrwa D (2006) Nanotechnology: assessments and perspectives, vol 27, Wissenschaftsethik und Technikfolgenbeurteilung. Springer, Berlin
- Schummer J (2007) Impact of nanotechnologies on developing countries. In: Allhoff F, Lin P, Moor J, Weckert J (eds) Nanotechnology: the ethical and social implications of nanotechnology. Wiley, Hoboken, pp 291–307
- Seaton A, Tran L, Aitken R, Donaldson K (2010) Nanoparticles, human health hazard and regulation. *J R Soc Interface* 7:S119–S129
- Selgelid M (2009) Dual-use research codes of conduct: lessons from the life sciences. *Nanoethics* 3:175–183
- Smalley R (2001) Of chemistry, love and nanobots. *Sci Am* 285:76–77
- Treder M, Phoenix C (2007) Challenges and pitfalls of exponential manufacturing. In: Allhoff F, Lin P, Moor J, Weckert J (eds) Nanoethics the ethical and social implications of nanotechnology. Wiley, Hoboken, pp 311–322
- Weckert J (2010) Nanotechnology, health and energy'. *Aust J Prof Appl Ethics* 11:45–55



# **8 Risk and Risk-Benefit Evaluations in Biomedical Research**

*Annette Rid*

Institute of Biomedical Ethics, University of Zurich, Zurich, Switzerland

<b>Introduction .....</b>	<b>180</b>
<b>History .....</b>	<b>181</b>
<b>Current Research .....</b>	<b>184</b>
Empirical Evidence on the Practice of Risk-Benefit Evaluations .....	185
Existing Ethical Frameworks for Risk-Benefit Evaluations .....	186
Component Analysis .....	187
Critical Appraisal of Component Analysis .....	188
The Integrative Approach .....	190
Critical Appraisal of the Integrative Approach .....	191
The Agreement Principle .....	194
Critical Appraisal of the Agreement Principle .....	194
Net Risks Test .....	196
Critical Appraisal of the Net Risks Test .....	198
Practical Implications of the Different Frameworks .....	199
The Debate About Upper Limits of Acceptable Research Risk .....	200
Risk Limits in Research without Informed Consent .....	200
Risk Limits in Research with Informed Consent .....	202
<b>Further Research .....</b>	<b>204</b>
What Fundamentally Justifies Exposing Research Participants to Risks? .....	204
When Does Research Have Social Value? .....	205
What Level or Type of Social Value Is Necessary to Justify Increasing Risks to Research Participants? .....	205
What Types of Potential Benefit for Participants Should Be Considered in Risk-Benefit Evaluations? .....	206
How Should We Define and Delineate Upper Limits of Acceptable Research Risk? ...	207
Is the Current System of Research Oversight Sufficiently Sensitive to the Risks Posed by the Research? .....	207

**Abstract:** One of the fundamental ethical concerns about biomedical research is that it exposes participants to risks for the benefit of others. Therefore, a key ethical requirement for biomedical research studies is that they have an acceptable risk-benefit profile. Yet, despite widespread endorsement of this requirement, how it should be implemented remains controversial. The present paper critically reviews recent debates about risk and risk-benefit evaluations in biomedical research. It traces the history of risk-benefit evaluations in research, which were traditionally conceived of as an extension of the risk-benefit assessment occurring in clinical care. From there, the paper presents and evaluates the four existing ethical frameworks for risk-benefit evaluations: the component analysis, the integrative approach, the agreement principle, and the net risks test. It is argued that the net risks test is superior to the alternative approaches, but fails to offer guidance for evaluating the ethical acceptability of risks that participants incur for research purposes only. This leaves two of the fundamental problems of risk-benefit evaluations in research inadequately addressed, namely, (1) how to weigh the risks to the individual research participant against the potential social value of the knowledge to be gained from a study and (2) how to set upper limits of acceptable research risk. Discussions about the “minimal” risk threshold in research with participants who cannot consent, such as children or patients with dementia, go some way to specifying upper risk limits in this study context. However, these discussions apply only to a small portion of research studies. The paper ends by highlighting several important questions that future research will need to address.

## Introduction

---

Progress in clinical care relies on the development of new preventive, diagnostic, therapeutic, and palliative methods. To determine whether potential new interventions represent an advance over current methods, they need to be tested in clinical research studies. Some such interventions might provide clinical benefit to the individuals who participate in these studies; however, most have no clinical effect and can even make participants worse off than they would have been outside the potential trial. For instance, many of the procedures necessary to evaluate the safety and efficacy of new interventions – such as blood draws, imaging procedures, or biopsies – offer essentially no clinical benefit to participants. The same is true in the early stages of drug development, which typically involve healthy volunteers who will not gain any clinical benefits. Similarly, though basic science research in biomedicine – for example, natural history studies or research in pathophysiology or - biomchemistry – is crucial for the development of new clinical interventions, it uses research interventions that offer participants no clinical benefits. As conventionally understood, the only benefit from performing these interventions is the generalizable knowledge intended to improve the care of future patients.

The conduct of biomedical research therefore serves as an example of one of the most fundamental quandaries in moral theory: when is it acceptable to expose individuals to some harm or risk of harm for the benefit of others? This question is not only of great theoretical interest, but also of tremendous practical importance. Conceptions of acceptable research risk influence both the extent to which participants are protected from risks of harm as well as the amount and type of research that is being conducted. A high threshold of acceptable research risk implies that most biomedical research will be carried out, but the risks to individual participants are likely to be excessive in at least some studies. By contrast, a very low threshold of acceptable risk will preclude a lot of research from being conducted, some of which is likely

to involve acceptable risks. This could unnecessarily restrict progress in clinical care and our understanding of disease. Risk-benefit evaluations in research therefore need to strike an appropriate balance between protecting participants from excessive risks, while allowing acceptable research to proceed.

To be ethically appropriate, biomedical research studies must satisfy a number of ethical requirements, including the requirement of a reasonable risk-benefit ratio (Emanuel et al. 2000). Most commentators, and essentially all research guidelines and regulations, agree that it is acceptable to expose research participants to some risks for the benefit of others, provided that (1) the risks to participants are minimized and reasonable in relation to the potential benefits for them and/or the potential social benefits gained from the generalizable knowledge produced by the study, and (2) the risks to participants who cannot consent are no more than “minimal” when the study offers no prospect of direct clinical benefit for them (Levine 1986; Emanuel et al. 2000; World Medical Association 2008). These requirements provide a general framework for risk-benefit evaluations. However, they offer little concrete guidance for how to make the judgments they mandate. Thirty years ago, the influential Belmont Report, which lays out the fundamental ethical principles for research involving human subjects, emphasized the “metaphorical” nature of most approaches to evaluating whether the risks and potential benefits of biomedical research studies are “balanced” or produce “a favorable ratio” (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research 1979). More than 20 years later, a similar document stated that then-current regulations uniformly mandate evaluation of research risks and benefits, but fail to explain how this should be done or make additional guidance available (National Bioethics Advisory Commission 2001). This verdict still holds true today.

Compared to the informed consent process for research participation – a topic that has generated a vast body of literature – risk-benefit evaluations in research remain an understudied area of inquiry. Nonetheless, scholars in bioethics have made significant progress in clarifying key questions and concepts for evaluating research risks and potential benefits over the past decade. For instance, several ethical frameworks for risk-benefit evaluations have now been proposed (Rajczi 2004; Weijer and Miller 2004; London 2007; Wendler and Miller 2007). Upper limits of acceptable research risk, especially in the context of pediatric research, have sparked significant debate (Freedman et al. 1993; Kopelman 2004; Resnik 2005; Wendler 2005; Ross and Nelson 2006). Commentators have also explored whether there should be any upper limits on the research risks to which competent, consenting adults may be exposed (Miller and Joffe 2009). The present paper will trace these recent debates. Although significant progress has been made, the paper argues that two of the fundamental questions of risk and risk-benefit evaluations in research remain inadequately addressed. First, how should we weigh the risks that the individual research participant assumes purely for research purposes against the potential social value of the knowledge to be gained from a study? Second, how can we justify and delineate upper limits of acceptable research risk? The entry ends by highlighting several important questions that future work will need to address.

## History

The acceptability of research risks has been a concern in the scientific community long before the current ethical and regulatory framework for biomedical research emerged in the second

half of the twentieth century. Investigators designed and conducted experiments with human subjects following longstanding moral traditions, which evolved from the early eighteenth century onward (Halpern 2004). By the early decades of the twentieth century, there was widespread agreement that research is ethical only when it rests on recognized scientific theories and is methodologically sound, given that scientifically invalid studies cannot advance medical knowledge or contribute to the collective good. Researchers were committed to testing medical innovations on animals before introducing them into clinical practice; researchers would also suspend or delay human applications if the risks of an intervention proved too serious in animal studies. In an attempt to reduce risks to participants, many investigators would try risk-laden interventions on themselves before embarking on tests with research participants. Moreover, the use of an investigational intervention was considered justified only when its risks were thought to be lower than the risks of the natural disease that the intervention was designed to prevent or treat. With regard to consent, investigators sharply distinguished between research with healthy subjects and research with patients as well as between therapeutic and nontherapeutic research. Consent was obtained in “nontherapeutic” research and in research with healthy subjects because the research offered no prospect of clinical benefit for participants. By contrast, when participants were also patients and the research intervention was expected to yield clinical benefits, investigators acted consistently with the reigning norms of clinical care, which did not require consent for providing beneficial treatments.

While the historical record attests to these moral traditions regarding risks in research participation, it also reveals tremendous inconsistency in whether and how this moral framework was implemented (Halpern 2004). This is in part due to the ambiguities inherent in the past traditions. For instance, the distinction between research and treatment as well as therapeutic and nontherapeutic research is often blurry, making it difficult to determine when consent ought to have been obtained. However, the inconsistency in implementing the traditional norms regarding research risk also traces back to investigators acting in violation of these norms. In many ways, the current system of independent oversight for research has evolved in reaction to a history of egregious abuse of research participants (Jonsen 1998).

The atrocious experiments with prisoners in Nazi concentration camps often resulted in the planned death of research subjects (e.g., high altitude experiments) and left countless subjects disfigured or permanently disabled (e.g., freezing experiments, sulfonamide experiments). Numerous studies involved such severe pain that they have been equated to medical torture. All experiments were conducted against the will of participants (Mitscherlich and Mielke 1949). The Nuremberg Code, which sets out ten ethical principles for the conduct of research, was written in response to these abuses as part of the allies’ verdict in the Nuremberg trials. Among other things, it states that the “voluntary consent of the human subject is absolutely essential” (Annas and Grodin 1992). In 1964, the World Medical Association followed suit and published the Declaration of Helsinki, which was a response to the Nazi experiments as well as a reaction to the Nuremberg Code (World Medical Association 1964). This document laid out a somewhat different set of ethical principles for biomedical research. Notably, the 1964 Declaration of Helsinki reintroduced the traditional distinction between therapeutic and nontherapeutic research and weakened the Nuremberg Code’s strict consent requirement for so-called therapeutic research. The Declaration also allowed research with participants who cannot give their own informed consent, provided that the consent of guardians or relatives is obtained.

Shortly after the Declaration of Helsinki was promulgated, a Harvard physician publicized a series of cases of abusive research performed in the USA (Beecher 1966). One of the listed studies, conducted by Drs. Chester Southam and Emanuel Mandel, involved the injection of cancer cells under the skin of elderly incapacitated patients. In 1966, the Southam study led the US National Institutes of Health (NIH) to require independent peer review for all of the research receiving NIH funding. This requirement built on prior practices of independent review for research occurring within the NIH Clinical Center, which focused on higher risk studies with healthy volunteers. External review was intended to ensure an independent risk-benefit evaluation, as well as adequate provisions for investigators to obtain the voluntary and informed consent of research participants (Jonsen 1998). The decision made by the NIH marks the establishment of the system of independent ethical and regulatory oversight for research that we know today.

The independent review of research protocols became a legal requirement for biomedical research – first in the USA and subsequently in most countries around the globe – after another research scandal shocked the American public. (However, the system of regulatory oversight in the USA is particular in the sense that research regulations pertain only to federally funded research, as well as to research data that are submitted to the Food and Drug Administration for obtaining market approval for new interventions.) In 1972, the infamous Tuskegee syphilis trial was publicized in the *New York Times*. The Tuskegee trial was a natural history study of syphilis that involved approximately 600 black men who were largely poor and uneducated. The study began in 1934 when syphilis was an incurable disease. However, it continued to follow the natural progression of syphilis even after penicillin became widely available as an effective treatment in the late 1940s. Trial participants were never told of their disease and never treated for it. They were also never informed that they were participating in a research study and that treatment for their condition could have been provided (Jonsen 1998). The Tuskegee study led to the establishment of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, the organization that wrote the influential Belmont Report (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research 1979). The Belmont Report enumerates the guiding ethical principles for the US Code of Federal Regulations for the protection of human subjects, which was first adopted in 1981 and served as a template for research regulations in many other countries. The Belmont Report also advanced the general ethical framework for research risks and potential benefits that remains widely recognized today (see  [Introduction](#)).

The recent history of research ethics and research oversight is insightful for the current state of the art regarding risk-benefit evaluations in at least two ways. First, one must remember that the current ethical framework for risk in research was largely developed in reaction to cases of blatant abuse and mistreatment of research participants. This probably helps to explain why the norms governing risk in research – and research more generally – can seem so “exceptional” when compared to the norms that govern risk in other areas of life, such as work or transport (Sachs 2010; Wertheimer 2010). Second, risk-benefit evaluations in biomedical research – and again, the ethics of research more generally – have a long tradition of being influenced by the norms that govern clinical care. Some of the clinical traditions, such as the distinction between therapeutic and nontherapeutic research and the application of clinical norms in so-called therapeutic studies, have now been widely rejected by commentators and guidelines – including the Declaration of Helsinki, which was revised several times since 1964 (Levine 1999; World Medical Association 2008). However, discussions continue about the distinction between

therapeutic and nontherapeutic research *interventions* and the appropriate role of “clinical equipoise” for evaluating the risk-benefit profile of therapeutic procedures (Weijer and Miller 2004; Wendler and Miller 2007). Much of this debate is fueled by a controversy about the clinical orientation of research ethics. This finding is unsurprising given that many participants in biomedical research are also patients and many investigators are also clinicians, which can give rise to dual roles and dual loyalties. However, it will become clear throughout this paper that the appeal to clinical norms is often inappropriate in the research context.

## Current Research

---

Before reviewing the recent debates about risk-benefit evaluations and upper limits of acceptable research risk, it is helpful to begin with several preliminary remarks regarding relevant concepts and terminology. First, although the term risk-benefit assessment or risk-benefit analysis is commonly used, it is strictly speaking a misnomer (Levine 1986). Risk-benefit evaluations are performed before the study begins; hence one cannot be certain what impact – harmful or beneficial – the study will have. Unlike the term “benefit,” however, the term “risk” captures the likelihood of harm. Therefore, a more precise terminology would be “risk-potential-benefit” evaluations of biomedical research. Because this expression is so unwieldy, the present entry will nonetheless use the term risk-benefit evaluations.

Second, because risk-benefit evaluations are performed before a study begins, the validity of these judgments must be determined *ex ante*. The fact that a study resulted in knowledge of important social value does not change the risk-benefit ratio in retrospect. As the physician-investigator Henry Beecher famously noted, “an experiment is ethical or not at its inception. It does not become ethical *post hoc*” (Beecher 1966). Conversely, the fact that a study seriously harmed some subjects does not necessarily imply that the study exposed participants to excessive risk (Emanuel and Miller 2007). Indeed, assuming we allow research that poses very low risks of serious harm, it follows that some serious harm eventually will occur in the context of ethically acceptable research.

Third, research risks and potential benefits are both a function of two more basic components: (1) the likelihood that a harmful or beneficial event or experience will occur as a result of a research intervention, and (2) the extent to which the event or experience, should it occur, sets back or advances the individual participant’s and/or society’s interests. For example, an allergy skin test might pose a risk of 13–26 per 100,000 performed tests of evoking a mild allergic reaction, which involves self-limiting hay fever symptoms or urticaria requiring antihistamine treatment. If these symptoms occur, they might be considered a small setback to a participant’s interests. Risk-benefit evaluations thus require *empirical judgment* about how likely it will be for different harms or benefits to result from the given research interventions within a study, as well as the study as a whole. Separate evaluation regarding how robust and relevant the available data are to make these judgments are also required. Risk-benefit evaluations also require *normative judgment* regarding the magnitude of the respective harms and benefits to participants, should they occur, and how much value the collected data would have for society at large.

It is widely agreed that risk-benefit evaluations should comprise judgments about all types and degrees of potential harm for participants, including physical, psychological, social, and economic harms (Levine 1986). However, commentators disagree as to whether all types of

potential benefit for participants should be included. Current ethical and legal guidance, as well as common wisdom in bioethics, focus exclusively on the potential *clinical* benefits of research interventions (Macklin 1989; King 2000). This might seem peculiar for readers without a medical background. However, the focus on potential clinical benefits is probably explained by the fact that risk-benefit evaluations in research have traditionally been conceived of as an extension of risk-benefit assessment in clinical care. In the clinical setting, the risks of a procedure need to be outweighed by the procedure's potential clinical benefits because clinicians are obliged to act in the patient's best clinical interests (hence the focus on potential clinical benefits). Recently, however, some commentators have argued that risk-benefit evaluations in research are distinct from risk-benefit evaluations in clinical care and thus the potential economic, social, or psychological benefits participants might realize during the study should also be factored into the risk-benefit calculus. Examples of these benefits might include payment, praise, or feelings of altruism (Jansen 2009; Sachs 2010; Wertheimer 2010); however, this view remains controversial (see [Further Research](#)).

Fourth, risk-benefit evaluations involve a set of judgments that go beyond a direct assessment and weighing of research risks and potential benefits. The term "risk-benefit evaluation," strictly understood, refers to absolute judgments about risks and potential benefits. However, risk-benefit evaluations in research have complex comparative and normative aspects. They typically start with a determination of whether the proposed study achieves the necessary minimum level of social value, given widespread endorsement of the view that research without any social value cannot justify exposing participants to any risks, however small (Emanuel et al. 2000). Efforts to reasonably reduce or minimize the risks to research participants and to enhance the potential benefits for them and society are generally seen as part of the evaluation process as well. Moreover, the risks and potential clinical benefits of so-called therapeutic research interventions need to be evaluated in comparison to the risks and benefits of available alternative treatments, if there are any available. Inclusion of these complex comparative and normative judgments into the process of evaluating research risks and potential benefits can seem confusing and might stand in the way of rigorous conceptual and normative analysis. However, including these judgments makes sense from a practical perspective. The existing ethical frameworks for risk-benefit evaluations are intended to serve as practicable guidance. Such guidance must ensure that investigators, sponsors, ethics committee members, and others systematically address all aspects related to evaluating the risks and potential benefits of biomedical research studies.

## Empirical Evidence on the Practice of Risk-Benefit Evaluations

---

In most countries, research ethics committees (RECs) are responsible for implementing the requirement that biomedical research studies have a reasonable risk-benefit ratio. However, surprisingly little research has investigated how RECs make these judgments. The few empirical studies that are available have produced concerning results. A semi-structured interview study from the Netherlands examined how REC members evaluate the risk-benefit profile of research protocols. In this study, only 6 of 53 reviewers indicated that they use a systematic approach; instead reviewers reported that they largely rely on their intuitive judgment (Van Luijn et al. 2002). While intuition plays an important role in risk-benefit evaluations, reliance on *mere* intuition – where one attempts to determine the risk-benefit profile of procedures and studies

based simply on how risky and beneficial they seem, without any appeal to intervening steps, analysis, or empirical data – increases the chances for mistakes. Extensive research from psychology shows that intuitive risk judgments are subject to systematic cognitive biases (Tversky and Kahneman 1974; Slovic 1987; Weinstein 1989). For example, people tend to judge familiar activities as less risky than unfamiliar activities. This bias increases the chances that those familiar with a research intervention will judge it to be low risk, while those not familiar with the intervention will judge it to be higher risk.

Indeed, several studies have found significant variation in how RECs evaluate the risks and potential benefits of research interventions and protocols. A follow-up study from the Netherlands found that 43 REC members varied greatly in their risk-benefit judgments about a vignette breast cancer study (van Luijn et al. 2006). Thirty percent believed that the risks of the study outweighed the benefits, 21% believed that the benefits outweighed the risks, and 35% assigned approximately equivalent weights to the study's risks and potential benefits. A survey of 188 REC chairpersons in the USA found a similar variation in risk judgments. Respondents were asked to categorize different research interventions based on the regulatory definition of minimal risk in the US Code of Federal Regulations (see [Risk Limits in Research Without Informed Consent](#)). To offer one example, 23% categorized allergy skin testing as minimal risk in healthy 11-year-olds, 43% categorized the same procedure in the same population as a minor increase over minimal risk, and 27% categorized it as more than a minor increase over minimal risk (Shah et al. 2004). A comparable study from Germany, which asked REC chairs to classify several research interventions as minimal or greater than minimal risk, reported similar findings (Lenk et al. 2004). Consistent with these findings, paradigm cases of minimal risk research were reviewed very differently by different RECs (Hirshon et al. 2002; McWilliams et al. 2003; Green et al. 2006a; Mansbach et al. 2007). It seems unlikely that the actual risks of research interventions, such as allergy skin testing, vary to this extent between research sites. This suggests that the current variation in risk-benefit evaluations is at least to some extent unjustified. Some RECs may be underestimating the risks posed by research interventions, thereby failing to protect participants from excessive risks. Other RECs may be overestimating the risks, thereby inadvertently blocking acceptable research. These findings expose the pressing need to develop practicable guidance for risk and risk-benefit evaluations in biomedical research.

## Existing Ethical Frameworks for Risk-Benefit Evaluations

---

Ever since research ethics became a field of inquiry in the mid-1960s, the following has been a key question: under which conditions are the risks to individual participants acceptable in light of the potential social benefits of the research? It is therefore surprising that scholars have only started to systematically address this question over the past decade. The following section critically reviews the existing four ethical frameworks for risk-benefit evaluations in biomedical research. The discussion will reveal that approaches to evaluating the risks and potential benefits of research studies have increasingly shed their clinical orientation. The first framework that was proposed, the component analysis, distinguishes between “therapeutic” and “nontherapeutic” research interventions and essentially requires that “therapeutic” interventions comply with the norms of clinical care. The integrative approach maintains the distinction between “therapeutic” and “nontherapeutic” procedures, but offers a different normative

justification for it. By contrast, the two remaining frameworks – the agreement principle and the net risks test – formulate the same ethical requirements for all research interventions. These approaches are guided by the idea that biomedical research is fundamentally different from clinical care and thus needs to satisfy its own set of ethical requirements. The four existing frameworks for risk-benefit evaluations differ primarily in how to evaluate the risks and potential benefits of so-called “therapeutic” research interventions, which offer participants a prospect of direct clinical benefit.

## Component Analysis

---

The first systematic approach to risk-benefit evaluations was developed by the physician and philosopher Charles Weijer in a paper commissioned, and subsequently endorsed, by the US National Bioethics Advisory Commission (Weijer 2001). Several other versions of this paper exist, as well as papers foreshadowing it (Weijer 1999, 2000, 2001; Weijer and Miller 2004; Miller and Weijer 2006; Freedman et al. 1992). However, the present discussion is primarily based on the commissioned paper. The approach is labeled “component analysis” because it requires independently evaluating the risks and potential benefits of each intervention or procedure involved in a research study, as opposed to making a global risk-benefit assessment of the study as a whole.

While the requirement to independently assess the risks and potential benefits of each research intervention is shared by most approaches to risk-benefit evaluations, component analysis has the distinctive feature of dividing the individual research interventions included in a study into two groups: therapeutic and nontherapeutic. Therapeutic research interventions are understood as those that are administered with therapeutic intent or warrant, such as investigational drugs. Nontherapeutic procedures are administered solely for the purpose of answering scientific questions. For example, blood draws, biopsies, or imaging procedures are often performed purely for research purposes in order to test the safety and/or efficacy of an investigational drug.

Component analysis takes the difference in intent or warrant to be morally significant. It stipulates that the risks of therapeutic interventions be justified only by the intervention’s potential clinical benefits for the individual research participant, whereas the risks of nontherapeutic interventions are justified by the knowledge gained from including the interventions in the study. Component analysis thus calls for “rigorous separate moral calculi” to evaluate the risks and potential benefits of therapeutic and nontherapeutic research interventions (Weijer 2001).

Therapeutic interventions must meet only one ethical requirement. Just like regular clinical procedures, therapeutic research procedures are ethically acceptable only if they promote participants’ clinical interests. To satisfy this requirement, the risk-benefit profile of therapeutic research interventions must be at least as favorable for the individual participant seen as the risk-benefit profile of the established standard of care, if one exists. Clinical equipoise is an indicator of when this requirement is met. Component analysis thus regards clinical equipoise as a (derivative) ethical requirement for therapeutic research interventions.

Clinical equipoise is commonly defined as the “honest, professional disagreement among expert clinicians” regarding which of one or more treatments or interventions is to be preferred from the point of view of the patient-participant (Freedman 1987a). It is assumed that

a disagreement of this sort will appear when the available evidence suggests that the interventions under consideration have comparable risk-benefit profiles. For example, the equipoise requirement allows randomization between an experimental and a standard treatment only when expert clinicians disagree about which of the treatments, and any other available treatments for the condition in question, has a more favorable risk-benefit profile. If one of the available interventions is clearly superior, the equipoise requirement – and hence component analysis – mandates that investigators provide the better treatment to all research participants. In this situation, component analysis precludes the conduct of the trial.

In contrast to therapeutic research interventions, component analysis requires that nontherapeutic interventions meet three distinct ethical requirements. First, the risks associated with nontherapeutic interventions must be minimized, to the extent that doing so is consistent with sound scientific design. Minimizing risks can involve avoiding unnecessary research procedures, identifying less risky methods to test a study hypothesis, or excluding participant groups who are at increased risk of being harmed. Second, the risks posed by nontherapeutic interventions must be reasonable in relation to the knowledge to be gained from the study. Third, in research involving vulnerable populations – commonly defined as research with participants who cannot give their own informed consent, such as children or patients with dementia – the risks of nontherapeutic interventions must be no more than a “minor increase” over the minimal risks posed by daily life activities. No upper limit of risk is specified for the use of nontherapeutic interventions in research with competent, consenting participants.

## Critical Appraisal of Component Analysis

---

Component analysis was the first systematic account of risk-benefit evaluations in biomedical research and thus a crucial step toward improving the unstructured approach used by most RECs. It rightly mandates the evaluation of individual research interventions, rather than making a global risk-benefit judgment about a research study as a whole. However, there are several fundamental problems with component analysis.

First, the distinction between therapeutic and nontherapeutic research interventions is not always clear (Wendler and Miller 2007). For example, an investigational drug in the earliest phase of drug testing is not administered with therapeutic intent because the goal is to gather preliminary data about drug safety, not efficacy. Administering so-called phase 1 drugs therefore seems to be a clear example of a nontherapeutic research intervention, especially as the trials in which they are used typically enroll healthy volunteers who do not stand to gain any clinical benefits from participating in a study. However, available data suggest that phase 1 drugs against cancer offer a prospect of direct clinical benefit for study participants (Miller and Joffe 2008). Thus, in some cases, it would not be unreasonable for investigators to offer these drugs with therapeutic intent or warrant. It is therefore unclear whether phase 1 cancer drugs are therapeutic or nontherapeutic interventions. Similarly, most research interventions offer participants at least some chance of potential clinical benefit. For example, CT scans of the brain that are performed solely for research purposes sometimes detect a treatable brain cancer (Yue et al. 1997; Katzman et al. 1999). Again, it is unclear whether these interventions qualify as therapeutic or nontherapeutic.

Second, even if the distinction between therapeutic and nontherapeutic interventions could be clarified, it does not seem morally relevant for evaluating the risks and potential benefits of research interventions (Wendler and Miller 2007). The goal of risk-benefit evaluations is to ensure that research interventions and studies do not expose participants to excessive risks of harm for the benefit of others. Whether a given level of research risk is excessive depends on the magnitude of the risks, the level of corresponding potential benefits for the participant, if any, and the level of potential social benefit from performing the intervention and study. It does not depend on whether the risks result from a therapeutic or a nontherapeutic intervention. Component analysis contends that “separate moral calculi” are necessary to prevent the risks of nontherapeutic interventions from being justified by the potential clinical benefits of therapeutic interventions (Weijer 2001). However, while this argument supports making a separate risk-benefit evaluation for each research intervention performed in a study, it does not support the use of different methods to make these evaluations. The differential ethical requirements for therapeutic and nontherapeutic research interventions therefore seem to lack a compelling normative justification (Wendler and Miller 2007).

Third, absent a compelling justification, component analysis leads to inconsistent risk-benefit evaluations (Wendler and Miller 2007). By formulating differential ethical requirements for therapeutic and nontherapeutic interventions, component analysis essentially introduces different thresholds of acceptable “net” risk in biomedical research (net research risks are risks that are not, or not entirely, offset by potential clinical benefits for participants; see  [The Net Risks Test](#)). Component analysis allows competent participants to consent to significant risks without any compensating potential clinical benefits as long as the risks result from a nontherapeutic intervention (e.g., a liver biopsy performed for research purposes only). There is no threshold for risks that result from nontherapeutic interventions provided that participants consent and the risks of the intervention have been minimized and are reasonable in relation to the social value of the study. By contrast, component analysis does *not* allow competent participants to consent to risks, even if they are outweighed by compensating potential clinical benefits for them, if the risks result from a therapeutic intervention that has a less favorable risk-benefit profile than available alternative treatments (e.g., a slightly less effective, first-generation cancer drug in a trial comparing the effectiveness of first- and second-generation treatments). This is due to the fact that therapeutic interventions must satisfy the requirement of clinical equipoise and hence cannot have even a slightly less favorable risk-benefit ratio than available alternative treatments. However, considering that the main goal of risk-benefit evaluations is to protect participants from being exposed to excessive risks for the benefit of others, it is inconsistent to allow significant net risks to competent participants for some research interventions, but not for others. Component analysis introduces a further inconsistency by – probably unintentionally – requiring that only the risks of nontherapeutic interventions, but not the risks of therapeutic procedures, be minimized.

Finally, the particular requirement of clinical equipoise for therapeutic interventions seems flawed from an ethical point of view. By adopting a (partially) clinical perspective on evaluating research risks and potential benefits, component analysis introduces the equipoise requirement to ensure that it is acceptable for clinicians to offer research enrollment to their patients. The underlying assumption is that physician-investigators, just like regular clinicians, have an obligation to act in the best clinical interests of research participants. Therefore, offering participation in research must be consistent with this obligation. Component analysis

contends that physician-investigators, by offering research enrollment, continue to act in their patients' best interests when there is a state of genuine uncertainty as to whether the experimental treatment or the standard treatment is superior (i.e., the risk-benefit profile of both treatments is seen as comparable).

The appeal to the ethical norms that govern clinical care in the context of research has been heavily criticized (Miller and Brody 2003). Central to this criticism is the observation that key elements of research – notably, research interventions lacking clinical benefits and the random assignment of different treatments (“randomization”) – are incompatible with a focus on the patient's best interests and individualized decision-making that is characteristic of clinical care (Miller and Brody 2003; Miller and Brody 2007). It has therefore been argued that clinical medicine is a fundamentally different activity than clinical research, and that the norms governing one practice do not apply to the other. To illustrate this point, it is helpful to contrast the fundamental ethical norms governing clinical care and biomedical research. Clinical care aims to benefit the individual patient. Hence, the central ethical principle governing this practice is that clinicians benefit their patients and not harm them. In contrast, biomedical research aims to benefit society by generating generalizable knowledge intended to help future patients. The central ethical principle governing research is that investigators should not exploit research participants by exposing them to excessive risks of harm for the benefits of others. Notably, this principle does not exclude duties of beneficence in research with sick patient-participants. However, the duty not to exploit participants is primary, and investigators' obligations of beneficence are constrained by achieving the scientific aims of a study (Miller and Brody 2007; Joffe and Miller 2008). Acknowledging this fundamental difference between the normative foundations of clinical care and biomedical research appears essential for developing an adequate framework for risk-benefit evaluations. Component analysis fails to do this.

## The Integrative Approach

The integrative approach was developed as an alternative framework for risk-benefit evaluations in biomedical research (London 2006, 2007). According to its author, philosopher Alex John London, traditional research ethics rests on one of two mistaken assumptions: (1) biomedical research is an inherently utilitarian endeavor and (2) moral constraints on the conduct of research are grounded in role-based obligations of either clinicians or researchers. The proposed alternative is to view biomedical research as a cooperative endeavor that is based on liberal-egalitarian ideals. On this view, research is one element within a larger social division of labor that is aimed at creating and maintaining social institutions that foster and advance the basic interests of each community member. The egalitarian component of this view is that each community member has a just claim to equal treatment regarding his or her basic interests. The liberal component is that a basic interest is defined as the fundamental interest in cultivating and exercising those human capacities that are necessary for developing a conception of the good, as well as formulating and pursuing a life plan based on this conception. Basic interests are contrasted with personal interests which individuals have given the particular conception of the good they have developed.

Based on this liberal-egalitarian perspective, the integrative approach sets out two requirements for research risks to be ethically acceptable. First, research risks must result from the least

amount of intrusion into subjects' personal and basic interests necessary for facilitating sound scientific inquiry. Second, research risks must be consistent with an equal regard for the basic interests of research participants as well as members of the larger community whose interests the research is intended to serve.

The second requirement is further specified by two "operational criteria" and a "practical test" to check the first criterion. The first operational criterion specifies the general requirement of equal regard for everyone's basic interests for the group of research participants whose basic interests are compromised by sickness, injury, or disease. Although this is not stated explicitly, the first operational criterion likely applies to research interventions that offer potential clinical benefits for participants. The criterion requires that "therapeutic" research interventions – borrowing the language of component analysis – must protect and advance the basic interests of research participants in a way that does not fall below the threshold of competent clinical care. The proposed practical test to delineate this threshold is to establish that expert clinicians are uncertain about the superiority of the given treatment options, including the investigational drug and/or placebo – a test that has great resemblance to the equipoise requirement in component analysis.

The second operational criterion specifies the requirement of equal regard for everyone's basic interests for those research interventions that pose risks to participants' basic interests without offering potential clinical benefits for them. This criterion requires that "nontherapeutic" interventions pose risks to the basic interests of the individual participant that are no greater than the risks to the basic interests that are accepted in the context of other socially sanctioned activities. To serve as appropriate comparators, these other activities must be similar in structure to research. The integrative approach sets out four necessary requirements for structural similarity: (1) the risks associated with the comparator activity do not add to the activity's social value (the risks are a "necessary evil"); (2) the activity is subject to active public oversight, which is intended to ensure that the risks posed by the activity have been deemed socially acceptable after due reflection; (3) individuals bear the risks of the activity primarily for the benefit of others; and (4) the activity must involve a principal–agent relationship in which one person (the agent) acts in the interests of another (the principal; in the research context, the investigator is seen as the agent and the participant as the principal).

## Critical Appraisal of the Integrative Approach

---

The integrative approach sets out to provide an alternative framework for risk-benefit evaluations in biomedical research. However, in terms of substantive normative content, its overlap with component analysis is striking. Once the complicated set of basic concepts, normative justifications, operational criteria, and practical tests is disentangled, the integrative approach formulates requirements that are highly reminiscent of those in component analysis. It is possible and helpful to recategorize the integrative approach into four ethical requirements that are needed for an acceptable risk-benefit ratio. First, the research must have social value. Second, the risks of all research interventions must be minimized. Third, the risk-benefit profile of research interventions that offer potential clinical benefits for participants – therapeutic procedures using the terminology of component analysis – must be at least as favorable as the risk-benefit profile of the established standard of care if one exists. To make this determination, a slightly modified version of the equipoise requirement is proposed. Fourth, the risks of research interventions that offer no

potential clinical benefits – or nontherapeutic procedures – must be no greater than those of socially sanctioned activities that are relevantly similar to research.

This recategorization shows that the basic setup of the integrative approach is very similar to component analysis. Both approaches distinguish between therapeutic and nontherapeutic research interventions and both require a state of clinical equipoise regarding the risk-benefit profile of therapeutic interventions. However, in contrast to component analysis, the integrative approach explicitly requires that risk-benefit evaluations begin with a judgment about the social value of the research. It also requires that the risks of all research interventions, not only nontherapeutic procedures, be minimized. Both additions are a significant improvement over component analysis. Furthermore, the integrative approach and component analysis advance different substantive limits for the risks of nontherapeutic interventions. Component analysis requires that the risks of nontherapeutic interventions must be outweighed by the potential social benefits of the research, and for research with vulnerable populations, the risks must be no greater than a minor increase over minimal risk. Hence, component analysis sets an upper risk limit for nontherapeutic interventions only when the interventions are performed in participants who cannot give their own informed consent. In contrast, the integrative approach limits the risks of nontherapeutic procedures for *all* participants – including competent, consenting participants – by requiring that the risks be no greater than the risks experienced in activities that are socially sanctioned and structurally similar to research.

The integrative approach is interesting for two reasons. First, of the existing ethical frameworks for risk-benefit evaluations, it is the only one that explicitly sets an upper limit of acceptable research risk for research with competent, consenting participants (see ➤ [Further Research](#)). The underlying idea is that limits to research risks should not be derived from utilitarian considerations. A utilitarian approach to risk-benefit evaluations would recognize risk limits only in cases where the study's potential benefits for participants and/or for society do not outweigh the risks. It follows that a study of tremendous social value could involve very high risks to participants. The integrative approach rejects this line of thought in favor of a “commitment to moral equality” (London 2006). In order to safeguard equality in the research context, the integrative approach permits research participants to assume risks to their basic interests only when the level of risk is no greater than the level of risk involved in other socially sanctioned activities that are structurally similar to research. Due to the absence of a general theory of acceptable research risk, risk comparisons are a common strategy for delineating upper risk limits. However, although such comparisons are the only viable approach at this point, risk comparisons are fraught with two fundamental problems. First, risk comparisons are circular to the extent that independent normative criteria for evaluating the risks of other activities are missing – there is no way of determining whether the risks of the comparator activity are acceptable. A specific problem for the integrative approach is that the risks of activities similar to research might be justified on utilitarian grounds, thus reintroducing a utilitarian thread that the framework wants to avoid. Second, it is difficult to determine when risks comparisons are valid. What exactly makes an activity relevantly similar to biomedical research? The integrative approach proposes four criteria, but further analysis is needed to determine whether these four criteria are both necessary and sufficient for establishing structural similarity.

The upper limits of acceptable risk as set out in the integrative approach pose a further problem. It is widely recognized that greater social value is necessary to justify higher risks to participants. However, the integrative approach merely requires that a given study have some

social value without exposing participants to greater risks than activities that are structurally similar to research. The integrative approach therefore fails to incorporate considerations of how a study's potential benefits for society should relate to the risks posed on individual subjects. This is problematic because studies of acceptable but relatively minor social value are permitted as long as the involved risks are not greater in other accepted and relevantly similar activities. Considering that volunteer fire fighting and paramedic services are the two comparator activities explored by London, the integrative approach likely leads to exposing participants to excessive risks in at least some research studies with relatively small potential social benefit.

Taking into account the relationship between individual risks and potential social benefits does not commit one to pursuing the greatest happiness for the greatest number of people, which is the utilitarian calculus that London rightly rejects. To the contrary, in order to protect participants from excessive risks, it is necessary to set upper risk limits that recognize a limited or constrained proportionality of the risks to participants and the potential social benefits for society. For research risks not to be excessive, greater social value is necessary to justify higher risks to participants – up to the independent upper limit of acceptable research risk that even tremendous social value cannot justify.

The second reason why the integrative approach is interesting is that it offers a new normative justification not only for the distinction between therapeutic and nontherapeutic research interventions, but also for the differential ethical requirements that these two classes of interventions allegedly need to meet. The integrative approach explicitly rejects role-based obligations as a normative foundation for risk-benefit evaluations and thus moves away from the (partial) clinical orientation of component analysis. Instead, the integrative approach is based on the idea that the basic interests of all members of a community require equal consideration. In the research context, this purportedly implies that (1) therapeutic research interventions must not undermine the competent clinical care of research participants and (2) nontherapeutic interventions must not compromise the basic interests of research participants to a greater extent than socially sanctioned activities that are structurally similar to research.

This liberal-egalitarian justification raises a number of questions. Most fundamentally, a liberal-egalitarian conception of social justice is only one of several competing theories of justice. Egalitarian theories of justice, with or without a liberal slant, have not been shown to be decisively superior to sufficientarian or prioritarian theories of justice. But even if one accepts its liberal-egalitarian foundation, the integrative approach continues to raise questions. Similar to component analysis, it fails to give a compelling justification for why nontherapeutic research interventions are allowed to pose risks to participants' basic interests, while therapeutic interventions are not. Indeed, since the integrative approach is fundamentally based on equal consideration for the basic interests of all community members, the differential regard for participants' basic interests seems especially difficult to justify. Furthermore, the integrative approach appears to consistently give priority to equality (i.e., equal consideration for the basic interests of all community members) over liberty (i.e., respect for the personal interests of the individuals of whom the community is comprised). However, it is not clear why equality should be given priority in this way. For example, some individuals might have personal interests in participating in research that poses higher risks to them than the risks involved in socially accepted and structurally similar activities, whatever those activities turn out to be. What is the justification for not allowing these individuals to pursue their personal interests out of concern for equal to everyone's basic interests?

It is equally unclear why the principle of equal consideration for basic interests should preclude some individuals from being exposed to more risks than others. Whether everyone should be exposed to the same level of risks at least partially depends on how many people there are in a community, how many valuable projects there are to pursue, how many people the different projects require, and how many people are willing to engage in the projects. For example, imagine a community whose members are engaged in numerous valuable activities involving low risk for the individuals involved in them, and two activities of incredible social value that pose relatively high risks to the involved individuals. The integrative approach seems to prohibit these latter two activities on the basis of equal consideration for the basic interests of all community members. Yet it seems that the principle of equal consideration for basic interests should not necessarily exclude riskier activities if all community members support these activities, given their important social value, and agree to a fair principle by which the involved individuals are chosen.

## The Agreement Principle

---

The general framework for risk-benefit evaluations endorsed by most research regulations and guidelines requires that the risks exposed to individual research participants are outweighed by the potential benefits for them and/or for society. Philosopher Alex Rajczi rejects this requirement and proposes the “agreement principle” as an alternative approach to evaluating research risks and potential benefits (Rajczi 2004). Unlike the integrative approach, which puts equality first, the agreement principle gives clear priority to the liberty of individuals. The agreement principle is based on the “limited voluntarist” idea that a liberal state must allow citizens to invite others to engage in risky activities as long as a competent and informed decision-maker would accept such offers. In the research setting, this implies that the risks and potential benefits of biomedical research studies should be evaluated from the perspective of a competent and informed decision-maker.

The agreement principle does not specify substantive conditions under which research risks are acceptable. Rather, it formulates three requirements for competent and informed decision-makers that RECs should use to evaluate the risk-benefit profile of research studies. First, research risks and potential benefits should be evaluated from the perspective of competent and informed decision-makers who have a set of values that is, in the very least, minimally consistent, stable, and affirmed as their own. Second, the decision-makers must be fully informed about the nature of the protocol in question. Third, they must reason clearly about whether to enter the protocol using those values. According to the agreement principle, a research study has an acceptable risk-benefit ratio if a competent and informed decision-maker who satisfies these three requirements would enroll in the study.

## Critical Appraisal of the Agreement Principle

---

The agreement principle emphasizes the liberal presumption that people should be free to do what they wish with their lives, provided that their actions are based on competent and informed decisions. At least in liberal democracies, this presumption is likely to be endorsed by many people. The agreement principle does not reduce the ethics of research to the

requirements of informed consent; rather, it explicitly maintains the idea of independent risk-benefit assessment by requiring that RECs prospectively evaluate the risks and potential benefits of research studies from the perspective of a competent, informed decision-maker. According to the agreement principle, because liberal political systems are generally “limited voluntarist” about government intervention, the risk-benefit evaluations should take a “limited voluntarist” approach as well.

This justification for evaluating research risks and potential benefits from the perspective of a competent and informed decision-maker differs distinctly from a more common paternalistic framework. Empirical data show that many otherwise competent prospective participants exhibit a variety of decisional defects – for instance, a lack of requisite scientific and clinical knowledge and a tendency to overestimate the potential benefits and underestimate the risks of research (often termed the “therapeutic misconception” (Appelbaum et al. 1987)). These defects make it difficult for prospective participants to make informed decisions about research enrollment. The decisional defects have led some commentators to argue that requiring a favorable risk-benefit ratio is necessary to ensure that participants’ interests are protected during the informed consent process (Miller and Wertheimer 2007). In contrast to this paternalistic position, the primary concern of the agreement principle is not to protect participants’ interests. Rather, the principle aims to maximize individual liberty by limiting state intervention in those risky activities in which a competent and informed decision-maker would choose to participate.

In order to implement the agreement principle, RECs will have to make generalizations about competent and informed decision-makers. REC deliberations are therefore likely to be guided by what the average, normal – or perhaps “reasonable” – decision-maker would choose. However, it is unlikely that the average, normal person will knowingly engage in research that involves a high likelihood of serious harm, such as severe and permanent disability or death, only to promote medical progress. For example, most healthy volunteers would presumably not participate in a smallpox vaccine study that involves a disease challenge (to test the vaccine’s efficacy of preventing infection) when no effective treatment exists and the mortality rate from smallpox is approximately 30%. Or, perhaps some might, but only if the development of a smallpox vaccine had enormous social value and promoted their own interests in some way. Similarly, most patients would probably not engage in very risky research if it offered little prospect of clinical benefit for them and the research was still in its exploratory stages. This implies that the average person would enroll in a study only if the risks of participating in the research are not excessive and they are outweighed by the study’s potential clinical and/or social benefits. In other words, “traditional” risk-benefit evaluations are still needed to implement the agreement principle because they are central to the decision-making of average, competent individuals.

One might therefore assume that the agreement principle adds little to existing accounts of risk-benefit evaluations because it effectively endorses the requirement of a favorable risk-benefit ratio without adjudicating conceptual or normative differences in interpreting this requirement. Yet, the agreement principle differs from the other approaches because it is consistent with at least two research scenarios that would be prohibited by alternative frameworks for risk-benefit evaluations as well as the general framework endorsed by most research guidelines and regulations (see [● Introduction](#)). The first scenario is that some competent and informed decision-makers might agree to participate in research even if the potential benefits do not outweigh the risks. The second scenario is that some competent and informed decision-makers might agree to participate in research if a favorable risk-benefit ratio cannot

be established because the risks and/or potential benefits associated with the research are uncertain. Rajczi himself admits that it is unlikely for decision-makers to be willing to participate in research with an unfavorable or uncertain risk-benefit ratio. In the unusual cases where a competent and informed person would want to participate in such research, it remains unclear why RECs should adopt the perspective of these unique people when their responsibility is to ensure that research risks are reasonable for the average, normal person.

More fundamentally, it is difficult to see why research with an unfavorable risk-benefit ratio merits defense. Why would anyone be interested in such research other than the few competent and informed decision-makers who might agree to take part? Research without a favorable risk-benefit ratio also raises concerns that go beyond those related to protecting the interests of research participants. Although the normative justification for requiring a favorable risk-benefit ratio remains surprisingly underexamined, it seems that the requirement is also grounded, among other things, in the need to protect the professional integrity of physician-investigators, maintain public confidence in the research endeavor, and ensure that societal gains in health and well-being are not won at the cost of exposing participants to excessive risks – even when participants give their consent. The agreement principle does not capture these considerations.

Finally, unlike alternative approaches to risk-benefit evaluations, the agreement principle provides no way to evaluate the risks and potential benefits of research with subjects who are not competent and informed, like children or patients with dementia. Given the importance of conducting research to promote the health of these often underserved populations, this is a fundamental problem.

## Net Risks Test

The “net risks test” was developed by David Wendler and Franklin G. Miller in response to Charles Weijer’s component analysis (Wendler and Miller 2007). Like component analysis, it focuses on evaluating the risks and potential benefits of individual research interventions rather than the entire research study. However, the net risks test’s normative foundation is fundamentally different from component analysis. Component analysis grounds risk-benefit evaluations of so-called therapeutic interventions in the norms governing clinical care; on the other hand, for nontherapeutic interventions, it sets out a research-specific requirement of proportionality between the risks of such interventions and the potential social benefits of the research. By contrast, the net risks test evaluates *all* research interventions based on the same requirements.

The net risks test is grounded in the principle of nonexploitation, which is taken to be the fundamental ethical principle governing biomedical research. The principle of nonexploitation requires that research participants not be exposed to excessive risks of harm *for the benefit of others*. The net risks test is designed to ensure this requirement is met. Significant parts of the net risks test are therefore dedicated to separating research risks that participants assume for their own potential benefit (such as the risks posed by investigational drugs, which are often – but not always – outweighed by the potential clinical benefits) from research risks that participants are exposed to solely to achieve the scientific goals of the study so that future patients can potentially benefit (e.g., the risks of a biopsy performed to test the safety of an investigational drug). These latter types of risk are labeled “net” research risks – risks of harm that are not, or not entirely, offset or outweighed by the potential clinical benefits for

participants. If these net research risks are justified, they are justified by the social value of the information to be gained from the study – whether or not they result from a “therapeutic” or a “nontherapeutic” research intervention (if one were to apply this distinction from component analysis).

The net risks test approach specifies four requirements for evaluating the risks and potential benefits of biomedical research studies. First, the risks of each research intervention in the study should be minimized, and each intervention’s potential clinical benefits should be enhanced. Satisfying this requirement, however, should not compromise achieving the scientific aims of a study.

Second, a research intervention should neither imply an excessive exposure to risk for the individual participant nor an excessive decrease in potential clinical benefit if the participant is also a patient and alternative treatments exist. Implementing this requirement involves assessing whether any research interventions in the study pose “net” risks to the individual participant and, in a second step, evaluating whether these net risks are excessive. Interventions that pose no net risks require no further evaluation given that the risks of these interventions are offset by the potential clinical benefits for the individual participant.

To determine whether a research intervention poses net risks, it is helpful to distinguish three types of net risks (Rid and Wendler 2011). *Absolute net risks* arise when the risks of an intervention are not outweighed by the intervention’s potential clinical benefits. Absolute net risks are *pure* when the research intervention poses risks without offering any potential clinical benefits for participants and *impure* when a research intervention offers potential clinical benefits for participants, but the benefits do not outweigh the risks. For example, a lumbar puncture poses impure, absolute net risks when it is performed primarily for research purposes – it offers some clinically relevant information that would not be available in routine clinical care, yet the benefits of this information typically will not be important enough to outweigh the risks of the lumbar puncture. (If the benefits did justify the risks, the procedure would likely be part of standard medical care.) *Relative net risks* arise when the risks of a research intervention are outweighed by its potential clinical benefits, but the intervention’s risk-benefit profile is likely to be less favorable than the risk-benefit profile of available alternative treatments or procedures. Finally, *indirect net risks* arise when a research intervention itself has a favorable risk-benefit profile, but the intervention diminishes the typical risk-benefit profile of other research or clinical procedures that are provided as part of, or in parallel to, the study. For example, an investigational drug against liver cancer might alter the risk-benefit profile of diuretics, a standard treatment for the ascites frequently associated with this cancer.

Once a research procedure is found to pose any of these three types of net risks to participants, the net risks test approach requires that the level of net risks be sufficiently low. To implement this requirement, it must be determined that the given intervention’s net risks fall below the absolute upper limit of acceptable net risks for research interventions. Arguably, there are levels of net risks that even tremendous social value and the informed consent of participants cannot justify (see [The Debate About Upper Limits of Acceptable Research Risk](#)). For example, one might think it unacceptable to expose research participants to a 30% risk of death for purely scientific purposes, even if participants consent. If the net risks of an intervention clearly exceed these absolute upper limits of acceptable research risk – whatever they turn out to be – the intervention should not be included in the study. If the scientific goals of the study cannot be achieved without including that intervention, then the study as a whole should be rejected.

Third, if a research intervention poses any net risks, these net risks must be justified by the expected knowledge gained from using that intervention in the study. This requirement ensures that net risks to the individual research participant – provided that they do not exceed the absolute upper limits of acceptable net risks – are not excessive in relation to the potential benefits to society.

Fourth, the *cumulative* net risks of all research interventions in a study must not be excessive. This requirement is intended to block research studies that include numerous procedures posing excessive aggregate net risks even though each procedure poses acceptable levels of net risks when evaluated separately.

## Critical Appraisal of the Net Risks Test

---

When compared to the alternative approaches to risk-benefit evaluations, the net risks test has the clear advantage of offering a consistent approach to evaluating research risks and potential benefits. The agreement principle leads to counterintuitive results because it endorses research with an unfavorable risk-benefit ratio and offers no guidance for research with participants who cannot consent. Neither component analysis nor the integrative approach gives a sound justification for why the risks of so-called therapeutic and nontherapeutic interventions should be held to different normative standards. The net risks test avoids these problems by requiring a favorable risk-benefit ratio both in research with competent, consenting participants and participants who cannot consent and by stipulating the same ethical requirements for all research interventions in a study, whether or not the interventions could be classified as “therapeutic” or “nontherapeutic.” A further advantage of the net risks test is that it captures the fundamental ethical concern about risks in research, namely, that study participants are exposed to risks for the benefit of others. It isolates the research risks that are not, or not entirely, offset by potential clinical benefits for participants, and evaluates these net risks in light of the social value of the research. By requiring that the net risks not be excessive, the net risks test approach also sets a clear limit to the risks that individuals are allowed to assume for the benefit of others (although it fails to clearly specify this limit; see below). The net risks test therefore explicitly rejects a utilitarian calculus of aggregating the net benefit of a study. At the same time, the test recognizes a constrained or limited proportionality of the net risks to individual participants and the potential social benefits of the research. This captures the common intuition that higher social value is necessary to justify higher risks to participants, while certain levels of research risk – for example, a very high likelihood of death – cannot be justified even by tremendous social value.

Despite these advantages, a major shortcoming of the net risks test is that it fails to specify several of the key concepts it invokes (London 2006, 2007; Weijer and Miller 2007). When are research risks “excessive”? To what extent is the answer to this question influenced by the study context (e.g., participants’ ability to consent)? Under which conditions are the net risks of a research intervention “justified” by the knowledge to be gained from the study? It would be unreasonable to expect any framework for risk-benefit evaluations to develop algorithms or decision rules to address these questions that would yield only one verdict about the risk-benefit profile of each possible protocol. Given the complexity of risk-benefit evaluations, any framework will have to rely on intuition and normative judgment, while clarifying and constraining the role played by intuition.

It is also important to note that upper limits of acceptable research risk and the appropriate relationship between individual risk and potential social benefit pose unmet challenges in research ethics. For example, much of the criticism raised against the net risks test equally applies to component analysis. Like the net risks test, component analysis offers no specific guidance for determining whether the risks of “nontherapeutic” interventions are reasonable in relation to the potential social benefits of the research. Component analysis uses the “risks of daily life” standard to set upper risk limits in research with vulnerable populations, but although this standard is less vague than the concept of “excessive” risk used by the net risks test, it is far from widely accepted (see [❸ Risk Limits in Research Without Informed Consent](#)). Finally, to fairly assess the net risks test – and any other framework for risk-benefit evaluations – it is important to distinguish between the formal method that one uses to evaluate research risks and potential benefits from the substantive standards that they must satisfy. The formal method guides the process of evaluating the relation between risks and potential benefits (e.g., component analysis turns around the distinction between therapeutic and nontherapeutic research interventions, whereas identifying any risks posed by solely for research purposes is central for the net risks test), and the substantive standards specify acceptable levels of research risk (e.g., component analysis invokes the “risks of daily life” standard for minimal risk in research with vulnerable populations). A complete approach to risk-benefit evaluations will require both: a correct formal method and appropriate substantive standards for delineating thresholds of acceptable research risk. The net risks test appears to offer the proper formal method, but it fails to specify the substantive standards that are also necessary for evaluating the risks and potential benefits of biomedical research studies.

## Practical Implications of the Different Frameworks

---

One might wonder whether the discussion of the existing frameworks for risk-benefit evaluations produces any practical relevance. While it is difficult to see how the agreement principle could be applied in practice, the remaining three approaches to evaluating research risks and potential benefits are sufficiently practical. Which approach is chosen will likely have significant practical implications. To illustrate this point, consider a study comparing a first-generation drug and a second-generation drug against liver cancer (drug A and B). Because drug A is the first-generation drug, it is highly likely that drug A will have a less favorable risk-benefit ratio than drug B. However, the two drugs have never been compared directly. Since A costs a fifth of what B costs, the study aims to determine whether B is sufficiently better than A to justify its higher costs. If judged by the requirements of component analysis and the integrative approach, the study would be rejected as ethically unacceptable for lacking clinical equipoise – it seems highly unlikely that expert clinicians would disagree regarding whether A or B is preferable from the point of view of prospective participants. By contrast, the study could be approved under the net risks test provided that (1) receiving drug A rather than drug B an excessive decrease in potential clinical benefit and/or excessive increase in risk (hence, it is not associated with excessive net risks) and (2) the net risks to subjects are justified by the value of the information to be gained from the study. Data regarding the comparative and cost-effectiveness of various treatments is vitally needed in order to effectively manage scarce health-care resources. A prohibition of comparative and cost-effectiveness research when the associated net risks to participants are

not excessive can undermine the goal of achieving or maintaining universal access to health care. The practical implications of endorsing the net risks test over component analysis or the integrative approach can therefore be consequential.

## The Debate About Upper Limits of Acceptable Research Risk

---

As mentioned above, how to define and delineate upper limits of acceptable research risk constitutes one of the unmet challenges in research ethics. This finding is unsurprising if one considers that risk limits in research pose one of the fundamental quandaries in moral theory – namely, when it is acceptable to expose individuals to some harm or risk of harm for the benefit of others. Some research ethicists have argued that it is impossible to answer this question because we lack a common measure or scale to compare research risks and potential benefits as well as determine their relationship to each other (Martin et al. 1995). However, while weighing individual risks against potential social benefits is difficult and raises profound philosophical questions (e.g., about the “separateness” of persons, Rawls 1971), it is important to note that these judgments are not unique to the research context. Indeed, most policy decisions depend on making trade-offs between individual risks and potential social benefits. For example, in deciding whether a new highway should be built near an elementary school, city planners must weigh the potential benefits to commuters against the risks posed to the children and set some upper limit of acceptable risk for the project. Although these judgments are complex and less than fully understood, we have clear normative ideals for how policy makers and public officials should make such decisions: they should serve as “ideal social arbiters” who (1) carefully consider the risks and potential benefits for all affected parties, ensuring that the risks to individuals are not excessive but rather proportionate to the personal and/or societal benefits; (2) give everyone’s claims fair consideration; and (3) treat like cases alike across different areas of policy (Rid and Wendler 2011). This suggests that trade-offs between individual risks and potential social benefits are not exceptional for biomedical research, and that they can be made, and often are made, based on clear – albeit very general – normative standards.

A recent paper argues that these standards can be fruitfully transferred to the research context (Rid and Wendler 2011). When considering upper limits of acceptable research risk in the given study context, reviewers should seek to adopt the perspective of an ideal social arbiter. Would a fully informed and impartial social arbiter recommend the study in question? To provide further guidance and context for these deliberations, different approaches to defining and delineating upper limits of acceptable research risk have been discussed in the literature. They are presented in the following two sections.

### Risk Limits in Research without Informed Consent

---

There is widespread agreement that research risks should be limited if the informed consent from participants is not obtained (e.g., in research involving deception) or cannot be obtained (e.g., in research involving children or incapacitated adults). Most ethical and regulatory guidelines define the upper limit of acceptable risk in this context as “minimal” risk, although a few documents allow for a “minor increase over minimal” risk in certain pediatric research studies (Department of Health and Human Services 1991; Ad hoc group for the development of implementing guidelines for Directive 2001/20/EC 2008).

Despite the widespread endorsement of the minimal risk threshold, there is no widely agreed definition of minimal risk. Different definitions have been proposed in the literature or implemented in regulatory documents. Some aim to specify the meaning of minimal risk, for example, by citing the lexical definition of minimal – the “least possible” risk (McIntosh et al. 2000) – or by stipulating that risks are minimal if they do not involve any risks of serious harm (Nicholson and Institute of Medical Ethics 1986) or result only in a slight and temporary health impact (Council of Europe 2005). Other definitions invoke risk comparisons, stating that research risks are minimal if they do not exceed the risks posed by routine clinical or psychological examinations (Council for International Organizations of Medical Sciences 2002; Kopelman 2004; Resnik 2005), daily life activities (Department of Health and Human Services 1991; Freedman et al. 1993), or charitable activities (Wendler 2005). Yet other definitions of minimal risk appeal to the moral responsibility of decision-makers. For example, some commentators have argued that risks are minimal if they would be endorsed by scrupulous parents or caregivers (Ackerman 1980; Nelson and Ross 2005; Ross and Nelson 2006).

It is beyond the scope of the present paper to critically review each of the proposed definitions of minimal risk. Readers should note, however, that all of the prominent definitions are problematic in some way. In an Additional Protocol to the Convention on Human Rights and Biomedicine, the Council of Europe requires that minimal risks in research result in no more than a “very slight and temporary” impact on participants’ health (Council of Europe 2005). However, a risk of serious or lasting harm can be “minimal” if the likelihood is sufficiently low. For example, a blood draw is widely – and arguably appropriately – considered a minimal risk procedure although it poses a very small risk of serious infection or permanent nerve damage. Definitions of minimal risk that strictly exclude risks of serious harm ignore the fact that research risks are a function of both the likelihood of a harm occurring and the magnitude of that harm, should it occur. Any risk threshold must therefore be sensitive to the relationship between these two basic components of risk.

The Council for International Organizations of Medical Sciences endorses the “routine examinations” standard for minimal risk, which requires minimal risks to be “no more likely and not greater than the risk attached to routine medical or psychological examination[s]” in the study population (Council for International Organizations of Medical Sciences 2002). However, some routine examinations clearly pose more than minimal risks, while certain nonroutine procedures pose essentially no risks. For example, a colonoscopy is a “routine” procedure in older adults as it is widely recommended by professional organizations and routinely performed (Levin et al. 2008). However, in addition to other risks, diagnostic colonoscopies are associated with a 30–650 risk of colon perforation per 100,000 performed procedures; such an outcome would typically require surgical management (Damore et al. 1996). This incidence of colon perforation likely exceeds the acceptable likelihood of research harms of this magnitude for minimal risk procedures. By contrast, Reiki – a Japanese technique where a clinician transfers healing energy by placing his or her hands a few centimeters away from the patient’s body – is not a routine procedure. However, it seems highly unlikely that Reiki poses any risks to patients.

Finally, the US Federal Regulations use the risks of daily life standard for minimal risk, which stipulates that “the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life” (Department of Health and Human Services 1991). However, many daily life risks – for instance, the risk of a child having an accident at home – are unacceptably high. It would

therefore be wrong to use daily life risks *per se* as a standard for acceptable research risk. Furthermore, the risks of daily life activities are dissimilar to the risks in research because individuals are usually exposed to them for their own benefit, rather than for the benefit of others (Kopelman 2004; Wendler 2005). For example, if someone takes the car to see his or her family or friends, he or she incurs the risks of the ride for the pleasure of seeing their loved ones. By contrast, although some research participants may experience personal satisfaction from helping others, they fundamentally assume the risks of research interventions to advance scientific understanding and/or the treatment of future patients. Risk comparisons can be useful when delineating upper limits of acceptable research risk (Rid et al. 2010), yet for these comparisons to be valid, it is crucial that the comparator activities be relevantly similar to research.

In addition to the conceptual and normative challenges of defining the minimal risk threshold, a further challenge is that there are very few methods for implementing the proposed definitions. Most guidelines and regulations use examples to illustrate the given definition of minimal risk. For example, the Office of Human Research Protections in the USA offers a list of minimal risk procedures used for collecting biological specimens. The list includes, among other interventions, non-disfiguring hair and nail clippings, the collection of external secretions, and the collection of mucosal cells by buccal swab (Office for Human Research Protections 1998). Illustrative examples are highly valuable. However, the available documents typically do not justify why the listed interventions qualify as minimal risk procedures. This is problematic because users cannot verify that the listed interventions indeed pose minimal risks except by relying on their intuitive judgment. Without a systematic method for making minimal risk judgments, it is also difficult to evaluate the risks of interventions that are not listed and to take into account the variation in risk depending on how a procedure is performed.

The only existing approach for implementing minimal risk definitions is the “systematic evaluations of research risks” (SERR) method (Rid et al. 2010). SERR allows REC members to implement risk definitions that invoke risk comparisons, setting out a method to systematically delineate, quantify, and compare the risks of research interventions with the risks of appropriate comparator activities. By explicitly drawing on the available data, and by evaluating the risks of research interventions in comparison to the risks of other activities, SERR has the advantage of reducing the influence of cognitive biases that can distort our intuitive judgments about risk (Tversky and Kahneman 1974; Slovic 1987; Weinstein 1989). However, the application of SERR remains limited at this point because the available data on the risks of research interventions are scarce (Rid and Wendler *forthcoming*), and we lack systematic criteria for identifying appropriate comparator activities (although the integrative approach offers some preliminary guidance; London 2006, 2007).

## Risk Limits in Research with Informed Consent

While there is widespread agreement that research risks should be limited if the informed consent from participants is not or cannot be obtained, most existing guidelines and regulations set no explicit upper risk limits in research with competent, consenting adults, provided that the risks are reasonable in relation to the potential social benefits of the research. The notable exception is the Nuremberg Code (Annas and Grodin 1992). The Nuremberg Code

stipulates that “no experiment should be conducted where there is an *a priori* reason to believe that death or disabling injury will occur; except, perhaps, in those experiments where the experimental physicians also serve as subjects.” This statement appropriately directs attention to the sometimes serious risks associated with research. However, it runs into a similar problem as those definitions of minimal risk which require that minimal risk procedures pose no risk of serious harm. Like these definitions, the Nuremberg Code ignores that limits of acceptable research risk need to take into account both the likelihood of a harm occurring and the magnitude of that harm, should it occur. The Code excludes research studies only if serious harm to participants is certain to occur (“...death or disabling injury *will* occur [emphasis added]...”). However, this risk limit seems too permissive when one considers that many people would probably consider a high likelihood of death or serious injury imposed for purely research purposes – for example, a 30–40% risk of death – to be unacceptable.

Although explicit regulatory guidance on absolute upper risk limits is largely absent, the vast majority of RECs are unlikely to approve research studies that involve a high risk of death or serious injury, even when the research involves competent, consenting adults. Moreover, the fact that the public has had strongly negative responses to the occurrence of severely adverse events in studies that offered no potential clinical benefits for participants – such as the death of Ellen Roche, a healthy volunteer in a study investigating the pathophysiology of asthma or the drastic immune response of several healthy subjects in the Phase 1 TeGenero trial (Steinbrook 2002; Suntharalingam et al. 2006) – suggests that many people endorse absolute upper limits of acceptable research risk. (One needs to consider, however, that these judgments may be distorted by hindsight bias). It also seems that no morally serious person would think it permissible to kill someone purely for research purposes, even if the investigator had obtained highly scrutinized voluntary and informed consent and the research had tremendous public health value. All this suggests that there are upper limits on acceptable research risk, even in the context of informed and voluntary consent. Yet, what defines these risk limits remains an open question.

The only academic paper dedicated to this issue argues that the uncertainty of producing potential social benefit from any particular research study calls for prudence in exposing individual participants to substantial research risks (Miller and Joffe 2009). Yet, this line of argument does not exclude exposing participants to very high risks if the social value of a given study is substantial and almost certain to materialize. However, this view goes against the widespread intuition that societal gains in health and well-being should not be won at the cost of exposing individual participants to excessive risks, even when they give highly scrutinized informed consent.

The integrative approach (London 2006, 2007) also sets upper risk limits in research with competent, consenting participants, although it does not explore the matter in detail. As discussed above, the approach requires that the risks of “nontherapeutic” research interventions be no greater than the risks of activities that are socially sanctioned and relevantly similar to research. Another example from the literature is live kidney donation (Miller and Joffe 2009). Although risk limits in research with competent, consenting participants remain an underexplored topic, a common thread from the debate surrounding minimal risk emerges in the attempt to define and delineate risk limits in comparison to the risks posed by other activities. This finding presses the need to develop criteria for identifying appropriate comparator activities for research.

## Further Research

### What Fundamentally Justifies Exposing Research Participants to Risks?

The current, general framework for risk-benefit evaluations puts great emphasis on consent as a condition of acceptable research risk. Only minimal risk research is permitted with participants who *cannot* consent, and it remains controversial whether even this research is justified (for a good overview in the context of pediatric research, see Wendler 2010). Conversely, there is no explicit upper limit of risk in research with participants who *can* consent, as long as the risks to participants are reasonable in relation to the potential benefits of the research (Miller and Joffe 2009). But is it the consent of participants that fundamentally justifies exposing participants to research risks? It is telling that commentators have addressed this question primarily in the context of research that involves participants who cannot themselves consent, such as children (McCormick 1976; Ackerman 1980; Freedman et al. 1993; Brock 1994; Ross 1998; Wendler 2010). There is no corresponding literature on research with competent, consenting participants; this seems to suggest that exposing participants to research risks is not – or significantly less – worrying when consent can be obtained. Moreover, those commentators who argue that research without consent is justified often invoke some approximation of consent. For example, advance consent (Ellis 1993), hypothetical consent (Maio 2002), presumed consent (Manning 2000), retrospective consent (Abramson et al. 1986), proxy consent (World Medical Association 2008), parental consent (Ross 1998), the obligation to consent (McCormick 1974), and hypothetical consent to the institution of clinical research (Brock 1994) have all been proposed to justify research with participants who are unable to make their own decisions.

Consent clearly has an important role to play. After all, most research interventions involve some level of intrusion into participants' privacy and/or bodily domain, which typically requires permission. But, unlike current approaches suggest, it appears that consent does not fundamentally justify research risks. If consent were fundamental, it would be either a sufficient or a necessary condition for acceptable research risk. Yet, as outlined above, most people would probably not think it permissible for an investigator to put study participants at a very high risk of death purely for research purposes, even if he or she obtains valid informed consent and the study has tremendous social value. Conversely, nearly everyone agrees that it may be permissible to conduct important research without consent, such as research on childhood vaccinations, when the risks are sufficiently low. This suggests that consent, while important, is neither sufficient nor necessary for acceptable risk in research.

Instead, it would seem that the relationship between risks to participants and the potential social value of research is fundamental for determining the acceptability of research risks. Without social value, investigators would not be justified in exposing participants to any research risks (Emanuel et al. 2000). This suggests that risk-benefit evaluations should fundamentally revolve around the relationship between the risks that individuals incur for purely research purposes and the potential social benefits of the research, with consent becoming a necessary, but not a sufficient condition for exposing participants to *substantial* risks of harm. This perspective seems promising because it would capture our intuitions both that consent is not sufficient to justify any level of research risk, and that consent is not always necessary in low risk research. However, to scrutinize these intuitions, and to develop a sound normative foundation for risk-benefit evaluations in research, we need more analysis of what

fundamentally justifies exposing study participants to research risks – those who cannot themselves consent as well as participants who can consent.

## When Does Research Have Social Value?

---

It is widely – although not universally – agreed that research must have social value for research risks to be acceptable (for authors who question that social value is a necessary condition for acceptable research risk, see Rajczi 2004; Sachs 2010; Wertheimer 2010). Despite this, there is no clear concept of what makes research socially valuable. According to some, scientific value is sufficient for social value (Freedman 1987b). To others, research is socially valuable when it has the potential to improve health and well-being (Levine 1986; Emanuel et al. 2000). However, if the potential for health improvements is unlikely and/or temporally distant – for example, in cellular pathology – this concept collapses into the concept of social value as scientific value. Moreover, it is not clear how the potential to improve health and well-being should be evaluated. Cost-effectiveness analysis, the dominant approach for evaluating social benefits in the allocation of scarce health-care resources, is notoriously blind to distributive concerns (Brock 2004). Moreover, traditional cost-effectiveness analysis has no means of factoring in the likelihood that a research study might lead to the successful development of a new intervention. The literature on research priority-setting has developed criteria to guide decisions about research funding that comprise social value judgments (Institute of Medicine 1998; Kaplan and Laing 2004; Council on Health Research for Development 2006; Hollis and Pogge 2008). Yet, it is unclear to what extent these criteria should be transferred to the context of evaluating the risks and potential benefits of research studies. For example, funding criteria might deliberately aim to encourage women to pursue a career in science. But although promoting gender equality in science is a socially valuable goal, gender equality is arguably not the type of social value that should offset risks to participants. More conceptual work on the social value of research is needed.

## What Level or Type of Social Value Is Necessary to Justify Increasing Risks to Research Participants?

---

Another unresolved question relates to the constrained or limited proportionality of individual risk and potential social benefit. There is widespread agreement that greater social value is necessary to justify higher research risks to participants until the absolute upper limit of acceptable research risk in the given population is reached. But when does one study have greater social value than another? Greater social value can be generated through more value of the same type (e.g., greater health improvement for more people as opposed to less improvement for fewer) or through social value of a different, more “valuable” type (e.g., research into an HIV vaccine of a foot fungus medication). One could imagine various ways in which different types or magnitudes of social value might be seen as justifying higher risks to participants, such as, clinical as opposed to “merely” scientific research, research aimed to address catastrophic threats rather than “ordinary” health issues, research that might benefit the worst-off instead of the well-off, “professional” research as opposed to research that is conducted as part of scientific training,

research to settle truly scientific rather than policy debates (or perhaps the other way around), and so forth. There is essentially no literature on what makes research more socially valuable such that the social value justifies exposing participants to increasing levels of research risk.

## What Types of Potential Benefit for Participants Should Be Considered in Risk-Benefit Evaluations?

---

Standard approaches to risk-benefit evaluations span all types of risk to participants, but they typically exclude certain types of potential benefits. Financial benefits – in particular, payment for participating in a research study – are usually excluded because they can turn an otherwise unfavorable risk-benefit ratio into a favorable ratio. The concern is that any study could have a favorable risk-benefit ratio for the individual participant as long as the payment is high enough (Macklin 1989). Although this is never stated clearly, the same line of argument would exclude other potential benefits for subjects from entering the risk-benefit calculus, for instance, gaining psychological satisfaction from contributing to progress in clinical care and science, accruing social esteem for one's contribution, and enjoying the interaction with investigators and other subjects. As mentioned above, the exclusive focus on the potential *clinical* benefits for participants is probably explained by the fact that risk-benefit evaluations in research have traditionally been conceived of as an extension of risk-benefit assessments in clinical care. However, this position has been questioned as part of the general move away from a clinical orientation in risk-benefit evaluations (Jansen 2009; Sachs 2010; Wertheimer 2010). In particular, commentators have argued that RECs should incorporate reasonable judgments about the financial payment when evaluating the risk-benefit profile of a study (Wertheimer 2010).

The argument for incorporating payment is typically based on two assumptions (Wertheimer 2010). First, people are paid in many other areas of life – in particular, the occupational sector – to incur risks they would otherwise not incur. There is no reason why research participants should not be treated the same. Second, research is special in that it is complicated and prospective participants suffer from widespread decisional impairments regarding their enrollment decisions. This pertains not only to the risks and potential clinical benefits of the research, but also to the level of payment necessary to fairly compensate for those risks: When one misjudges the risk or risk-benefit profile of a study, it is difficult to evaluate whether the level of compensation is appropriate. If one assumes that risk-benefit evaluations are justified on paternalistic grounds – as a way of protecting people from their own misjudgment (Miller and Wertheimer 2007) – it follows that payment should be incorporated into risk-benefit evaluations. Thus, the question of which types of potential benefits should be considered is closely linked to another unsolved problem – namely, what fundamentally justifies risk-benefit evaluations in research. Is it the need to protect participants' interests? Or are there reasons beyond paternalism, such as the need to protect the professional integrity of physician-investigators, to maintain public confidence in the research endeavor (Miller and Joffe 2009) and ensure that societal gains in health and well-being are not won at the cost of exposing even competent, consenting participants to excessive risks?

## How Should We Define and Delineate Upper Limits of Acceptable Research Risk?

---

As discussed above, one of the unsolved challenges in risk-benefit evaluations is the definition and delineation of upper limits of acceptable research risk. Perhaps the most dominant approach is to define upper risk limits in comparison to the risks of other activities. Commentators have invoked the risks of routine clinical or psychological examinations (Kopelman 2004; Resnik 2005), daily life activities (Freedman et al. 1993), and charitable activities (Wendler 2005) to define the minimal risk threshold in research without informed consent. Others have used volunteer firefighting, emergency assistance (London 2006, 2007), and live kidney donation (Miller and Joffe 2009) as possible comparator activities to delineate upper risk limits in research with competent, consenting participants. Given the absence of a general theory of acceptable risk – or more specifically, a general theory of acceptable research risk – it makes sense to set upper risk limits in comparison to the risks posed by non-research activities (see ➤ [Critical Appraisal of the Integrative Approach](#)). Although this approach is ultimately circular, it has important advantages. Risk comparisons provide a context for delineating risk limits, they promote consistent risk judgments across activities in different realms of life, and they allow policy makers and RECs to appeal to considered risk judgments made outside the research context (Rid et al. 2010). However, for risk comparisons to be valid, one must choose comparator activities that are relevantly similar to research and widely considered to be acceptable for the given population. Future research needs to specify both requirements.

## Is the Current System of Research Oversight Sufficiently Sensitive to the Risks Posed by the Research?

---

Recently, there has been increasing dissatisfaction and concern that the current ethical and regulatory framework for research does not adequately calibrate the level of safeguards for participants to the level of risk posed by a study. REC members complain that they spend too much time reviewing low-risk research, which diverts time and resources from the review of research that poses greater risks (Fost and Levine 2007; Kim et al. 2009). Investigators complain about the bureaucratic burden associated with REC review when the risks of the research are low which seems to have little to do with the protection of participants (O'Herrin et al. 2004; Gunsalus et al. 2006; Fost and Levine 2007; van Teijlingen et al. 2008). On the other hand, there is significant concern – often after the serious injury or death of research participants – that the current system of research oversight fails to protect participants from excessive risks (Savulescu et al. 1996; Savulescu 2002). Many therefore call for a “risk-based” system of research oversight that matches various safeguards for participants – including independent ethical review and safety monitoring – to the level of risk posed by the research (European Commission and Health and Consumers Directorate-General 2011; The Academy of Medical Sciences 2011). The idea of a “risk-based” system of research oversight makes intuitive normative sense. If the goal is to protect research participants from excessive risks of harm, more safeguards are necessary if the risks of the research are high, and less if the risks are low. However, it remains unclear what such a system would or should look like. Future research should therefore develop a normative framework for risk-based approaches to research oversight.

Readers will note that the listed questions for future research remain within the traditional focus of risk-benefit evaluations, which concentrates on weighing the risks of a study to participants against the potential benefits to them and/or society. However, when adopting a broader perspective on risk-benefit evaluations, further unresolved questions emerge. What should be done when research results risk being abused (e.g., research on bioterrorism) or the research has the potential to produce “negative” social value (e.g., research on the relation between intelligence and race; for a discussion of such “dual use” research, see Green 2006b; Selgelid 2007, 2009)? When the conduct of research exposes third parties to risks? For example, in challenge studies with infectious diseases, how should we address the risks to individuals who do not participate in the research (Kimmelman 2005; Resnik and Sharp 2006; Hausman 2007)? Risk-benefit evaluations in biomedical research raise a multitude of interesting and important questions for future research to address.

## Acknowledgments

This chapter is based to a significant extent on two papers that I coauthored with David Wendler, Ph.D., from the Department of Bioethics in the Clinical Center of the National Institutes of Health (Rid and Wendler 2010, 2011). My work on this paper was supported by the Swiss National Science Foundation.

Many thanks to Frank Miller for critical comments on an earlier draft of this chapter, and to Greer Donley for editing the manuscript.

## References

- Abramson NS, Meisel A, Safar P (1986) Deferred consent. A new approach for resuscitation research on comatose patients. *JAMA* 255:2466–2471
- Ackerman TF (1980) Moral duties of parents and nontherapeutic clinical research procedures involving children. *Bioethics Q* 2:94–111
- Ad hoc group for the development of implementing guidelines for Directive 2001/20/EC relating to good clinical practice in the conduct of clinical trials on medicinal products for human use (2008) European Union. Ethical considerations for clinical trials on medicinal products conducted with the paediatric population. *Eur J Health Law* 15:223–250
- Annas G, Grodin M (1992) The Nazi doctors and the Nuremberg code. Oxford University Press, New York
- Appelbaum PS, Roth LH, Lidz CW, Benson P, Winslade W (1987) False hopes and best data: consent to research and the therapeutic misconception. *Hastings Cent Rep* 17:20–24
- Beecher HK (1966) Ethics and clinical research. *N Engl J Med* 274:1354–1360
- Brock DW (1994) Ethical issues in exposing children to risks in research. In: Grodin M, Glantz L (eds) Children as research subjects: science, ethics, and law. Oxford University Press, New York
- Brock DW (2004) Ethical issues in the use of cost effectiveness analysis for the prioritisation of health care resources. In: Anand S, Peter F, Sen A (eds) Public health, ethics, and equity. Oxford University Press, New York, pp 201–223
- Council for International Organizations of Medical Sciences (CIOMS) (2002) International ethical guidelines for biomedical research involving human subjects. CIOMS, Geneva
- Council of Europe (2005) Additional protocol to the convention on human rights and biomedicine, concerning biomedical research. Council of Europe, Strasbourg, France
- Council on Health Research for Development (COHRED) (2006) Priority setting for health research: toward a management process for low and middle income countries. COHRED, Geneva
- Damore LJ 2nd, Rantis PC, Vernava AM 3rd, Longo WE (1996) Colonoscopic perforations. Etiology, diagnosis, and management. *Dis Colon Rectum* 39: 1308–1314

- Department of Health and Human Services (1991) U.S. Code of Federal Regulations, Title 45, Part 46
- Emanuel EJ, Miller FG (2007) Money and distorted ethical judgments about research: ethical assessment of the TeGenero TGN1412. *Am J Bioeth* 7:76–81
- Emanuel EJ, Wendler D, Grady C (2000) What makes clinical research ethical? *JAMA* 283:2701–2711
- European Commission, Health and Consumers Directorate-General (2011) Revision of the ‘Clinical Trials Directive’ 2001/20/EC. Concept paper submitted for public consultation
- Fost N, Levine RJ (2007) The dysregulation of human subjects research. *JAMA* 298:2196–2198
- Freedman B (1987a) Equipoise and the ethics of clinical research. *N Engl J Med* 317:141–145
- Freedman B (1987b) Scientific value and validity as ethical requirements for research: a proposed explication. *IRB* 9:7–10
- Freedman B, Fuks A, Weijer C (1992) Demarcating research and treatment: a systematic approach for the analysis of the ethics of clinical research. *Clin Res* 40:653–660
- Freedman B, Fuks A, Weijer C (1993) *In loco parentis.* Minimal risk as an ethical threshold for research upon children. *Hastings Cent Rep* 23:13–19
- Green LA, Lowery JC, Kowalski CP, Wyszewianski L (2006a) Impact of institutional review board practice variation on observational health services research. *Health Serv Res* 41:214–230
- Green SK, Taub S, Morin K, Higginson D (2006b) Guidelines to prevent malevolent use of biomedical research. *Camb Q Healthc Ethics* 15:432–439 (discussion 439–447)
- Gunsalus CK, Bruner EM, Burbules NC, Dash L, Finkin M, Goldberg JP, Greenough WT, Miller GA, Pratt MG (2006) Mission creep in the IRB world. *Science* 312:1441
- Halpern SA (2004) Lesser harms. The morality of risk in medical research. University of Chicago Press, Chicago/London
- Hausman DM (2007) Third-party risks in research: should IRBs address them? *IRB* 29:1–5
- Hirshon JM, Krugman SD, Witting MD, Furuno JP, Limcangco MR, Perisse AR, Rasch EK (2002) Variability in institutional review board assessment of minimal-risk research. *Acad Emerg Med* 9:1417–1420
- Hollis A, Pogge T (2008) The health impact fund: making new medicines accessible to all. Incentives for Global Health. [http://www.yale.edu/macmillan/igh/hif\\_book.pdf](http://www.yale.edu/macmillan/igh/hif_book.pdf)
- Institute of Medicine, Committee on the NIH Research Priority-Setting Process (1998) Scientific opportunities and public needs: improving priority setting and public input at the national institutes of health. National Academy Press, Washington, DC
- Jansen LA (2009) The ethics of altruism in clinical research. *Hastings Cent Rep* 39:26–36
- Joffe S, Miller FG (2008) Bench to bedside: mapping the moral terrain of clinical research. *Hastings Cent Rep* 38:30–42
- Jonsen AR (1998) Experiments perilous: the ethics of research with human subjects. In: Jonsen AR. The birth of bioethics. Oxford University Press, New York/Oxford, pp 126–165
- Kaplan W, Laing R (2004) Priority medicines for Europe and the world. World Health Organization, Geneva
- Katzman GL, Dagher AP, Patronas NJ (1999) Incidental findings on brain magnetic resonance imaging from 1000 asymptomatic volunteers. *JAMA* 282:36–39
- Kim S, Ubel P, De Vries R (2009) Pruning the regulatory tree. *Nature* 457:534–535
- Kimmelman J (2005) Medical research, risk, and bystanders. *IRB* 27:1–6
- King NM (2000) Defining and describing benefit appropriately in clinical trials. *J Law Med Ethics* 28:332–343
- Kopelman LM (2004) Minimal risk as an international ethical standard in research. *J Med Philos* 29:351–378
- Lenk C, Radenbach K, Dahl M, Wiesemann C (2004) Non-therapeutic research with minors: how do chairpersons of German research ethics committees decide? *J Med Ethics* 30:85–87
- Levin B, Lieberman DA, McFarland B, Andrews KS, Brooks D, Bond J, Dash C, Giardiello FM, Glick S, Johnson D, Johnson CD, Levin TR, Pickhardt PJ, Rex DK, Smith RA, Thorson A, Winawer SJ (2008) Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US multi-society task force on colorectal cancer, and the American College of Radiology. *Gastroenterology* 134:1570–1595
- Levine RJ (1986) Ethics and regulation of clinical research. Urban & Schwarzenberg, Baltimore
- Levine RJ (1999) The need to revise the declaration of Helsinki. *N Engl J Med* 341:531–534
- London AJ (2006) Reasonable risks in clinical research: a critique and a proposal for the integrative approach. *Stat Med* 25:2869–2885
- London AJ (2007) Two dogmas of research ethics and the integrative approach to human-subjects research. *J Med Philos* 32:99–116
- Macklin R (1989) The paradoxical case of payment as benefit to research subjects. *IRB* 11:1–3
- Maio G (2002) Ethik der Forschung am Menschen. Frommann-Holzboog, Stuttgart-Bad Cannstadt
- Manning DJ (2000) Presumed consent in emergency neonatal research. *J Med Ethics* 26:249–253
- Mansbach J, Acholonu U, Clark S, Camargo CA Jr (2007) Variation in institutional review board responses to

- a standard, observational, pediatric research protocol. *Acad Emerg Med* 14:377–380
- Martin DK, Meslin EM, Kohut N, Singer PA (1995) The incommensurability of research risks and benefits: practical help for research ethics committees. *IRB* 17:8–10
- McCormick RA (1974) Proxy consent in the experimentation situation. *Perspect Biol Med* 18:2–20
- McCormick RA (1976) Experimentation in children: sharing in sociality. *Hastings Cent Rep* 6:41–46
- McIntosh N, Bates P, Brykczynska G, Dunstan G, Goldman A, Harvey D, Larcher V, McCrae D, McKinnon A, Patton M, Saunders J, Shelley P (2000) Guidelines for the ethical conduct of medical research involving children. Royal College of Paediatrics and Child health: Ethics Advisory Committee. *Arch Dis Child* 82:177–182
- McWilliams R, Hoover-Fong J, Hamosh A, Beck S, Beaty T, Cutting G (2003) Problematic variation in local institutional review of a multicenter genetic epidemiology study. *JAMA* 290:360–366
- Miller FG, Brody H (2003) A critique of clinical equipoise. Therapeutic misconception in the ethics of clinical trials. *Hastings Cent Rep* 33:19–28
- Miller FG, Brody H (2007) Clinical equipoise and the incoherence of research ethics. *J Med Philos* 32:151–165
- Miller FG, Joffe S (2008) Benefit in phase 1 oncology trials: therapeutic misconception or reasonable treatment option? *Clin Trials* 5:617–623
- Miller FG, Joffe S (2009) Limits to research risks. *J Med Ethics* 35:445–449
- Miller FG, Wertheimer A (2007) Facing up to paternalism in research ethics. *Hastings Cent Rep* 37:24–34
- Miller PB, Weijer C (2006) Trust based obligations of the state and physician-researchers to patient-subjects. *J Med Ethics* 32:542–547
- Mitscherlich A, Mielke F (1949) Doctors of infamy: the story of the Nazi medical crimes. Henry Schuman, New York
- National Bioethics Advisory Commission (2001) Ethical and policy issues in research involving human participants. The Commission, Bethesda, MD
- National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1979) The Belmont Report. Ethical principles and guidelines for the protection of human subjects of research. U.S. Government Printing Office, Washington, DC
- Nelson RM, Ross LF (2005) In defense of a single standard of research risk for all children. *J Pediatr* 147:565–566
- Nicholson RH, Institute of Medical Ethics (Great Britain) (1986) Medical research with children: ethics, law, and practice: the report of an Institute of Medical Ethics working group on the ethics of clinical research investigations on children. Oxford University Press, Oxford, New York
- Office for Human Research Protections (OHRP) (1998) Categories of research that may be reviewed by the institutional review board (IRB) through an expedited review procedure
- Office for Protection from Research Risks (OPRR) (1993) Informed consent: legally effective and prospectively obtained. *Office of Protection from Research Risks*
- O'Herrin JK, Fost N, Kudsk KA (2004) Health insurance portability accountability act (HIPAA) regulations: effect on medical record research. *Ann Surg* 239: 772–776 (discussion 776–778)
- Rajczi A (2004) Making risk-benefit assessments of medical research protocols. *J Law Med Ethics* 32:338–348, 192
- Rawls J (1971) Theory of justice. Belknap Press of Harvard University Press, Cambridge, MA
- Resnik DB (2005) Eliminating the daily life risks standard from the definition of minimal risk. *J Med Ethics* 31:35–38
- Resnik DB, Sharp RR (2006) Protecting third parties in human subjects research. *IRB* 28:1–7
- Rid A, Wendler D (2010) Risk-benefit assessment in medical research – critical review and open questions. *Law, Probability Risk* 9:151–177
- Rid A, Wendler D (2011) A framework for risk-benefit evaluations in biomedical research. *Kennedy Inst Ethics J* 21:141–179
- Rid A, Wendler D (forthcoming) A proposal and prototype for a research risk repository to improve the protection of research participants. *Clin Trials*
- Rid A, Emanuel EJ, Wendler D (2010) Evaluating the risks of clinical research. *JAMA* 304:1472–1479
- Ross LF (1998) Children, families, and health care decision making. Clarendon Press, Oxford, New York
- Ross LF, Nelson RM (2006) Pediatric research and the federal minimal risk standard. *JAMA* 295:759 (author reply 759–760)
- Sachs B (2010) The exceptional ethics of the investigator-subject relationship. *J Med Philos* 35:64–80
- Savulescu J (2002) Two deaths and two lessons: is it time to review the structure and function of research ethics committees? *J Med Ethics* 28:1–2
- Savulescu J, Chalmers I, Blunt J (1996) Are research ethics committees behaving unethically? Some suggestions for improving performance and accountability. *BMJ* 313:1390–1393
- Selgelid MJ (2007) A tale of two studies; ethics, bioterrorism, and the censorship of science. *Hastings Cent Rep* 37:35–43
- Selgelid MJ (2009) Governance of dual-use research: an ethical dilemma. *Bull World Health Organ* 87:720–723

- Risk and Risk-Benefit Evaluations in Biomedical Research
- Shah S, Whittle A, Wilfond B, Gensler G, Wendler D (2004) How do institutional review boards apply the federal risk and benefit standards for pediatric research? *JAMA* 291:476–482
- Slovic P (1987) Perception of risk. *Science* 236:280–285
- Steinbrook R (2002) Protecting research subjects – the crisis at Johns Hopkins. *N Engl J Med* 346:716–720
- Suntharalingam G, Perry MR, Ward S, Brett SJ, Castello-Cortes A, Brunner MD, Panoskaltsis N (2006) Cytokine storm in a phase 1 trial of the Anti CD28 monoclonal antibody TGN1412. *N Engl J Med* 355:1018–1028
- The Academy of Medical Sciences (2011) A new pathway for the regulation and governance of health research. The Academy of Medical Sciences, London
- Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. *Science* 185:1124–1131
- Van Luijn HE, Musschenga AW, Keus RB, Robinson WM, Aaronson NK (2002) Assessment of the risk/benefit ratio of phase II cancer clinical trials by Institutional Review Board (IRB) members. *Ann Oncol* 13:1307–1313
- van Luijn HE, Aaronson NK, Keus RB, Musschenga AW (2006) The evaluation of the risks and benefits of phase II cancer clinical trials by Institutional Review Board (IRB) members: a case study. *J Med Ethics* 32:170–176
- van Teijlingen ER, Douglas F, Torrance N (2008) Clinical governance and research ethics as barriers to UK low-risk population-based health research? *BMC Public Health* 8:396
- Weijer C (1999) Thinking clearly about research risk: implications of the work of Benjamin Freedman. *IRB* 21:1–5
- Weijer C (2000) The ethical analysis of risk. *J Law Med Ethics* 28:344–361
- Weijer C (2001) The ethical analysis of risks and potential benefits in human subjects research: history, theory, and implications for US regulation. In: National Bioethics Advisory Commission (eds) Report on ethical and policy issues in research involving human participants. Volume II - commissioned papers and staff analysis, Bethesda, MD, p P1-P29
- Weijer C, Miller PB (2004) When are research risks reasonable in relation to anticipated benefits? *Nat Med* 10:570–573
- Weijer C, Miller PB (2007) Refuting the net risks test: a response to Wendler and Miller's "Assessing research risks systematically". *J Med Ethics* 33:487–490
- Weinstein N (1989) Optimistic biases about personal risks. *Science* 246:1232–1233
- Wendler D (2005) Protecting subjects who cannot give consent: toward a better standard for "minimal" risks. *Hastings Cent Rep* 35:37–43
- Wendler D (2010) The ethics of pediatric research. Oxford University Press, New York
- Wendler D, Miller FG (2007) Assessing research risks systematically: the net risks test. *J Med Ethics* 33:481–486
- Wertheimer A (2010) Rethinking the ethics of clinical research: widening the lens. Oxford University Press, New York
- World Medical Association (WMA) (1964) Declaration of Helsinki. In: Schmidt U, Frewer A (eds) History and theory of human experimentation: the Declaration of Helsinki and modern medical ethics. Franz Steiner Verlag, Stuttgart, pp 339–340
- World Medical Association (WMA) (2008) WMA Declaration of Helsinki – ethical principles for medical research involving human subjects (first version adopted 1964)
- Yue NC, Longstreth WT Jr, Elster AD, Jungreis CA, O'Leary DH, Poirier VC (1997) Clinically serious abnormalities found incidentally at MR imaging of the brain: data from the cardiovascular health study. *Radiology* 202:41–46



# 9 Understanding and Governing Public Health Risks by Modeling

Erika Mansnerus

London School of Economics, London, UK

<b>Introduction: Governing Public Health Risks .....</b>	<b>214</b>
About the Case Study .....	216
<b>Toward Modeled Encounters with Public Health Risks .....</b>	<b>217</b>
Current Modeling Techniques .....	219
<b>Predicting Infectious Risks Through Modeling .....</b>	<b>221</b>
Explanation-Based Predictions .....	222
Case of an Integrated Simulation Model of <i>Haemophilus influenzae</i> Type B (Hib)	
Transmission .....	223
“What Would Happen If” Questions as a Key to Explanation-Based Predictions .....	224
Building Pandemic Scenarios .....	226
Predicting the Pandemic .....	226
What Kinds of Models Are Used as Scientific Evidence Base for Preparedness Planning? .....	227
Fingerprint of the Pathogen, and of the Population .....	228
Behavioral Assumptions and Their Alternatives .....	228
Scenario-Building Predictions .....	229
Beneficial Encounters with Infectious Risks .....	230
<b>Further Research: Toward the Analytics of Risk .....</b>	<b>231</b>

**Abstract:** Increase in the use and development of computational tools to govern public health risks invites us to study their benefits and limitations. To analyze how risk is perceived and expressed through these tools is relevant to risk theory. This chapter clarifies the different concepts of risk, contrasting especially the mathematically expressed ones with culturally informed notions, which address a broader view on risk. I will suggest that a fruitful way to contextualize computational tools, such mathematical models in risk assessment is “analytics of risk,” which ties together the technological, epistemological, and political dimensions of the process of governance of risk. I will clarify the development of mathematical modeling techniques through their use in infectious disease epidemiology. Epidemiological modeling functions as a form of “risk calculation,” which provides predictions of the infectious outbreak in question. These calculations help direct and design preventive actions toward the health outcomes of populations. This chapter analyzes two cases in which modeling methods are used for explanation-based and scenario-building predictions in order to anticipate the risks of infections caused by *Haemophilus influenzae* type b bacteria and A(H1N1) pandemic influenza virus. I will address an interesting tension that arises when model-based estimates exemplify the population-level reasoning of public health risks but has restricted capacity to address risks on individual level. Analyzing this tension will lead to a fuller account to understand the benefits and limitations of computational tools in the governance of public health risks.

## Introduction: Governing Public Health Risks

---

- ▶ Nothing is a risk in itself; there is no risk in reality. But on the other hand, anything can be a risk; it all depends on how one analyses the danger, considers the event (Ewald 1991, p. 199).

In June 2009, we faced a risk of a global pandemic caused by A(H1N1) “swine flu” virus. At the moment WHO defined the outbreak a *pandemic*, a viral flu infection turned into a global risk. Two years later the risk assessment process that led to the global call is questioned. Was the “swine flu” a global risk after all? “Nothing is a risk in itself” shows that what is identified as a risk depends on the way in which “danger is analyzed” or how risks are identified.

What, then, is a risk? Risk research answers to that question through three main paradigms: the statistical-probabilistic, the epidemiological, and the sociological. Initially developed for insurance industry, the statistical-probabilistic approach applies various estimates of personal benefits, effectiveness of treatments, and use “risk calculations.” This means translating “successful calculations of risk” into an objective measure. Estimates of how various health-related factors influence the probability of falling ill is a typical example of epidemiological paradigm. It comprises of psychometric studies, which focus on understanding public risk preferences and applies mental modeling to analyze rational decision making. When cultural and individual perception and responses to risk are interlinked, sociological approach is in use. Within this sociocultural paradigm, risk is regarded as a “socially constructed phenomenon although it has some roots in nature.” Furthermore, we can identify governmentality approach, which is based on Foucault’s analysis of societal governance and includes into the analysis broader issues of power as a part of the sociological approaches to risk. This form of analysis includes “construction of realities through practice and sense-making, encompassing the multitude of societal organizations and institutions producing social reality” (Taylor-Gooby and Zinn 2006, p. 43).

In the public health domain, the idea of what counts as a risk has a dynamic nature. Risk from the public health policy point of view may not be a risk for an individual. Dean argues that epidemiological risk forms a “long-standing and pervasive form of risk rationality.” In his words, “epidemiological risk is concerned with the rates of morbidity and mortality among populations.” When talking about epidemiological risk, “the health outcomes of populations are subject of risk calculation” (Dean 2010, p. 218). In regard to public health risks, we can talk about risk rationality as a form of rationality, a way of thinking about and representing events, which happens through calculations (Dean 2010, p. 213). Castell (1991) shows how mental health problems shifted in their classification as dangerous to risk as a part of historical, theoretical, and practical shift toward *risk rationality*. Underneath the emphasis of population as a subject of risk calculation is the historical shift from a family to population as a re-centering concept of economy. Michel Foucault argues that “[...] population has its own regularities, its won rate of deaths and diseases. [...] [S]tatistics shows also that the domain of population involves a range of intrinsic, aggregate effects, phenomena that are irreducible to those of the family, such as epidemics. Population comes to appear above all else as the ultimate end of government” (Foucault 1978/1991, pp. 99–100). Population becomes the object of governance. The attitude or mentality of governance is known as governmentality in Foucault’s work. It refers to the forms of thought, expertise, and knowledge that direct and guide the acts of governance (Dean 2010). How the risk rationality that shifts “dangerousness” of disease outbreaks into risks manifest in epidemiology?

New risks “violate many assumptions of risk calculation,” as Taylor-Gooby and Zinn (2006, p. 25) argue. They are global, complex, and entangled with different areas. They share characteristics of catastrophes. These new risks are “mainly invisible and inaccessible by direct means.” They challenge the statistical-probabilistic approach to risk. How do we encounter these new risks? One way of coming into terms with them is suggested in Smith’s analysis of SARS (Sudden Acute Respiratory Syndrome) epidemic (Smith 2006, p. 3114). He proposes a mediatory approach in order to overcome the dichotomy between the realist and constructivist accounts to risk. According to his “material-discursive” position, “risk is both a materially measurable probability of an event and a socially constructed element of how that probability is perceived by the individual and society.” This mediatory approach is useful when explaining the ways in which risk is represented, anticipated, and processed in models. The mediatory approach does not solely lean on realist interpretation, which sees risk as an objective threat or danger that can be measured independently of the social context within which it occurs. Nor does it reduce risk to culturally or socially constructed threat, which cannot be demonstrated independently of those processes. Close to the realist interpretation is Schlich and Tröchler’s (2006) definition of risk and uncertainty. He says that “one can speak of risk when the probability estimates of an event are known or at least knowable, while uncertainty, by contrast, implies that these probabilities are inestimable or unknown.” Riesch (2011) provides a classification of uncertainties that problematizes the clear-cut division between probability estimates and unknown events. I suggest that new risks, such as emerging infectious outbreaks, can be encountered by accommodating statistical-probabilistic approach of risk calculation (in the form of mathematical modeling) with the questions of governance. In this chapter, I will apply the Foucauldian notion of governmentality and address risk calculation as a *technical rationality*. How could we see public health risks through this lens?

The heterogeneity of the definitions of risk suggests that understanding risk encompasses the following aspects: dangerousness of an event, unpredictability of its occurrence and course,

and severity of its consequences. Infectious disease outbreaks, therefore, form a source of “danger” that affects the public. This means that public health risks are anticipated through surveillance and monitoring procedures, which are carried out by the national public health institutes. Surveillance and monitoring procedures comprise of keeping records of notifiable diseases or participating in international collaboration to govern outbreaks of emerging infections. Surveillance activities are carried out on various levels: on national level by public health institutes (e.g., Health Protection Agency in the UK) and on international and intergovernmental level (e.g., by WHO, ECDC, the European Disease Control Centre). However, risks from infectious outbreaks create a challenge for public health decision makers, who aim at identifying the risks through preparedness planning and revising protective interventions, such as vaccinations. Their interest is to employ predictions of the course of the outbreak. In order to do that, decision makers search for alternative ways to process the information flow. Evidence for developing the required preventive and protective measures for decision-making processes is produced by computer-based modeling techniques. Hence, modeling techniques can be utilized in the encounters with public health risks from infectious diseases. These *modeled encounters* provide predictions that facilitate risk assessment processes. This chapter shows how two modes of prediction: explanation-based and scenario-building provide strategies, not only to produce evidence for the decision-making, but also to translate the potential threat to a quantifiable, measurable risk. By doing so, modeled encounters with risks allow us to follow the social processes that try to control and minimize public health risk in society. In his commentary on climate models, Hulme et al. (2009, p. 127) highlights an important aspect of models, which is highly applicable to predictions from epidemiological models: “Scientists and decision-makers should treat climate models not as truth machines, but instead as one of a range of tools to explore future possibilities.” Through the analysis of the two types of predictions, I will address how epidemiological models function as *technologies* when encountering public health risks.

How is understanding of public health risks formed, estimated, and communicated through modeling? How does public health risk prevention use technical understanding gained by model-based predictions? These are the main questions addressed in the analysis of *modeled encounters* with infectious risks, which arise from *Haemophilus influenzae* type b bacteria and from A(H1N1) pandemic influenzae (“Swine flu” outbreak in 2009). By analyzing these two cases, I will show how modeling provides a way to *encounter* risks, which means that modeling itself forms a base for risk calculation and estimation that allows rendering the available information into predictions (cf. Mansnerus 2009a, 2011).

## About the Case Study

The case study analyzed in this chapter was conducted during 2002–2004 at the University of Helsinki. I observed modeling practices in 22 work meetings (recorded and transcribed, duration of a meeting app. 2 h) at the National Public Health Institute (currently the Institute for Health and Welfare) and conducted 28 thematic, semi-structured interviews (transcribed for analysis) with mathematical modelers, epidemiologists, and computer scientists working as members of the interdisciplinary team (published in Mattila 2006a, b). The models were published in Auranen 1999, 2000; Auranen et al. 1996, 1999, 2000 and Auranen et al. 2004; Leino et al. 2000, 2002, 2004 and Mäkelä et al. 2003. The study analyzes how an integrated simulation

model on Hib transmission in the Finnish population produces *explanation-based* predictions. In this chapter, I will keep the focus on a single, integrated model in order to allow a detailed description of the ways in which the model predicts. The findings from the Hib case will be discussed in relation to microsimulation model on mitigating an *influenzae* pandemic. This example will show how microsimulations produce *scenario-building* predictions. The analysis focuses on detailed micro-level observations and interpretations of the predictive capacities of microsimulations. Both models are chosen, because they provide clear examples of the predictive capacities of simulations. I have chosen not to explore the vast literature on *pandemic influenzae* models, but to concentrate on a detailed level on a single model. The analysis is informed by a practical course “Introduction to Infectious Disease Modeling,” organized by the London School for Hygiene and Tropical Medicine, 2006, which gave me ability to read the models and understand their core structure. As a part of the coursework, we analyzed the published pandemic simulation models and prepared a group exercise on national preparedness planning. I have chosen one of these as an example of scenario-building modeling exercise.

The structure of this chapter is as follows: section [● Toward Modeled Encounters with Public Health Risks](#) discusses the development of modeled encounters in epidemiology. Section [● Predicting Infectious Risks Through Modeling](#) shows how this development takes place when mathematical models are used in public health decision making. I will use two cases as examples to show how public health risks can be governed through modeling. Section [● Further Research: Toward the Analytics of Risk](#) discusses how the tension between individual level risk perception and population-level risk assessment could be reconciled analyzing public health risks suggests further research within the analytics of government approach.

## Toward Modeled Encounters with Public Health Risks

---

Technologies form an integral part of the procedures through which organizations and individuals try to control risks they encounter. These technologies range from software systems to visualizations and representations, from advanced technological structures (e.g., air traffic control) to models (Hutter and Power 2005). Models, or broadly speaking computer-based tools and techniques have become commonly used in various scientific and policy-making contexts. Yet, they have the potential to “legitimate a range of possible social futures,” as Evans (2000) frames the capacity of economic models. Den Butter and Morgan (2000) seem to suggest that models, which are engaged with policy-making processes in economics, actually build a bridge between research and policies or between “positive theory and normative practice.” In their account, these models form a part of the “value chain” through which knowledge is created, stored, and transmitted in organizations.

One of the main reasons to develop modeling techniques is to overcome uncertainties related to complex phenomena, such as climate, economy, or infectious diseases. Establishing modeling practices also helps forming a network that integrates available knowledge and communicates it further. Paul Edwards (1999, p. 439) argues that “Uncertainties exist not only because quantifiable, reducible empirical and computational limits, but also because of unquantifiable, irreducible epistemological limits related to inductive reasoning and modeling”(cf. also Hillerbrand 2010). His argument seems to suggest that due to the very nature of the modeled phenomenon itself, uncertainties remain as a part of the process. As Shackley and Wynne (1996, p. 276) emphasize, scientific knowledge, or the “authority” of

it, is limited in policy making, since it prevents decisions to be made or actions taken. Van den Bogaard (1999, p. 323) shows, on the contrary, that the first macroeconomic model, developed by Tinbergen in the 1930s for the Dutch Central Planning Bureau, was a “liberation both from the uncertainties caused by the whimsical nature of the economy and the woolly theories of the economists.” One form of uncertainties remains within the models, as MacKenzie’s (2005, p. 186) study on financial economics shows, “models affect the reality they analyze.” According to him, this “reflexive connection serves to increase the veracity of finance theory’s assumptions and the accuracy of its models’ predictions,” but it may also function in a counterproductive way (as in his case, the exploitation of arbitrage opportunities by using mathematical models leads to instability of the system). It seems to me that when modeling complex, open systems, such as climate (cf. Gramelsberger 2010) or ecological systems, there will remain uncertainties, because of limited computational capacity, biased reflection of the reality, or unpredictable nature of the phenomena themselves.

Encounters with risk, as Hutter and Power (2005, p. 11) clarify, are events of problematization that “place in question existing attention to risk and its modes of identification, recognition and definition.” “Risk identification,” they continue, “is socially organized by a wide variety of institutions which support prediction and related forms of intervention around the possibility of future events.” When we encounter risks and uncertainties, or predict a possible course of events, we develop and utilize various measurement devices, such as statistical methods, surveys, and models. From a historical perspective, we can identify a shift away from “informal expert judgment toward a greater reliance in *quantifiable objects*,” as Porter (2000, p. 226) argues in his case study of the use of mortality statistics in life insurance industry. The tendencies underlying this shift are addressed by a growing interest in sociology of quantification – i.e., in the “production and communication of numbers.” How do we “do things with numbers?” Espeland and Stevens elaborate J.L. Austin’s idea of speech acts (doing things with words) to the domain of quantification and they call it “doing things with numbers.” They argue that as with words, “numbers often change as they travel across time and social space” (Espeland and Stevens 2008, pp. 402–406). The “change in numbers” could be seen as a parallel process to the one that characterizes how public health risks become quantified through its historical development. This historical development can be aligned with two processes: First, the development and application of mathematical methods in order to understand the dynamics of disease transmission (Mansnerus 2009a), and second, the shift within biopolitics (politics that is concerned with governance of living conditions in a population) toward risk politics (Rose 2001).

The first process, gaining understanding of disease transmission and developing tools to express and represent that process in mathematical terms arose when germ theory of disease located the cause of infections to their microbiological origins, germs. What initiated the move toward mathematical formulations were population-level observations of infectious cycles, such as influenza outbreaks in households in London 1890–1905, as Hamer documented (Hamer 1906). Later on, Kermack and McKendrick (1927) divided the population into specific subgroups, *compartments* of susceptible, infected and immune, which represented different phases observed during an epidemic outbreak. These developments in mathematical epidemiology aimed at identifying various factors that caused transmission of germs and spread of the infectious outbreak.

The second process, which developed toward risk politics, was grounded on the developments on the microbiological level, but emphasizes the ways in which concern for the health of

the population adopted preventive measures. In his analysis of the birth of social medicine, Foucault looks into the organization of the Health Service and the Health Office in England at the end of the nineteenth century. He shows how three functions developed:

- “Control of vaccination, obliging the different elements of the populations to be immunized.
- Organizing the record of epidemics and diseases capable of turning into an epidemic, making the reporting of dangerous illnesses mandatory.
- Localization of unhealthy places and, if necessary, destruction of those seedbeds of insalubrity” (Rabinow and Rose 1994, 335).

The first two aspects in the development of health services show how public health measures take the form of governance. These forms, namely “control of vaccination,” especially if understood as a process of assessing and revising vaccination schemes, and “organizing the record of epidemics” are present when modeling techniques are applied to public health risks. “Control of vaccination” is one component in modeling process; it is applied as a preventive measure, as an estimate in terms of “herd immunity,” immunity cover for the whole of population when only a significant portion of it is vaccinated. This can be obtained as an indirect observation from the models and it allows estimating the vaccination coverage needed to protect the whole population. Organizing the record of epidemics could be extended to cover the predictive functions, as the following case studies will show. So mathematical methods in epidemiology developed initially in conjunction with the early observations of infectious cycles and outbreaks that gave rise to develop preventive actions against these risks.

Even though these aspects are to some extent present in the preventive public health work, the intention behind infection prevention has changed. Rose (2001) argues that the shift toward risk politics happened when public health programs and preventive medicine were transformed and health became “economized,” meaning that individuals were expected to become active in maintaining their well-being and health. Whereas the earlier programs understood health as fitness and were hence framed to tackle the “unfitness of populations,” the current emphasis is on costs of ill health for the economy (Rose 2001, pp. 5–7). This shift results in various strategies for the government of risk. Rose says that risk denotes in this context “a family of ways of thinking and acting, involving calculations about probable futures in the present followed by interventions into the present in order to control that potential future.” And, he continues, demand for these collective measures increases. As we will learn from the detailed study of how modeled encounters with public health risks happen, I would argue that the shift toward risk governance is still partially embedded in the preventive ideals of public health risk perception. The predictions from the models I will study are not estimates of the economic costs of a pandemic, although those have been taken into account through different analyses. Model-based predictions seem to function as a way to estimate the need to vaccinate and to assess the spread of the infection. On the basis of these predictions, protective measures toward the public can be initiated. But how do we actually build models to predict public health risks?

## Current Modeling Techniques

Modeling techniques provide a way to produce predictions, or in broader terms evidence for decision-making processes and, as such, they are a new way to encounter public health risks.

Modeled encounters with public health risks are approached in terms of studying the nature of model-based predictions. These predictions form the core of our attempts to control public health concerns, prepare for sudden outbreaks, or estimate population-wide effects of bacterial or viral transmissions. Modeled encounters with risk introduce two modes of predictions, those based on explanations and those building scenarios for future events or developments. Scenario in this context means an outline of an imagined, possible situation that has been quantified through modeling. By locating modeling into the context of measurement, we will learn the different ways in which trust, credibility, and usability of the modeled predictions emerge and are communicated from research domain to decision-making processes (cf. Morgan and Morrison 1999; Boumans 1999; van den Bogaard 1999).

What do we then understand by modeling? Generally speaking, computer-based models (including simulation techniques) in infectious disease epidemiology share the following characteristics: First, they have a three-part elementary structure, which comprises of data element, mathematical method and computational techniques, and element of substantial knowledge, or epidemiological component. Secondly, they are “tailor-made,” usually addressing specified research question, which to some extent limits their applicability. Thirdly, majority of these models rely on currently available data. And it is precisely the need to reuse and reanalyze the data that partially motivate the model-building exercise. Fourthly, micro-practices that are independent the context of application, say the pathogen studied, can be identified within modeling process. A detailed analysis of the eight consecutive steps in modeling process is documented in Habbema et al. (1996, p. 167):

- Identification of questions to be addressed
- Investigation of existing knowledge
- Model design
- Model quantification
- Model validation
- Prediction and optimization
- Decision making
- Transfer of simulation program

The importance of setting the question follows the idea of *tailoring* a model to address particular interests. Investigation of existing knowledge is a process in which existing literature, laboratory results, experiences of existing models, and data from surveillance programs are integrated as a part of model assumptions. Morgan (2002) aligns model building with similar steps to those mentioned by Habbema et al. (1996), although her focus is on economic models. The main difference is that in her account the model is first to build to represent the world, then subjected to questions and manipulation in order to receive the answers to the questions, then relating the answer to real-world phenomena.

Model design follows the existing understanding of how the phenomenon of interest behaves and is often represented through a compartmental structure. Compartmental structure means that the population is divided into subgroups according to the impact on immunity, susceptibility, and potential recovery from the modeled infection.

Model quantification is the process of estimating the optimal parameter values and setting the algorithms to run the simulations. In Habbema's et al. (1996) account, model validation means checking the model against data from control program. The particular interest in this

chapter is to analyze how the step from prediction and optimization to decision making is taken in regard to public health risks.

By transfer of simulation program, Habbema et al. refer to the generalizability of the computer program in other infectious diseases. This step-wise characterization of the micro-practices of modeling highlight that modeling is an *iterative* practice, which builds upon and checks back with previous steps throughout the process. Importantly, these models are not only scientific exercises to develop better computational algorithms, they are built first and foremost to explain, understand, and predict the infectious disease outbreaks or transmission processes. The major application of this group of models (including also simulations) is to design, for example, reliable and cost-effective vaccination strategies or to predict the course of influenza pandemic (Mattila 2006a, b, c). Morgan (2001) characterizes this process as “storytelling,” in which a model is a narrative device. I suggest that scenario-building predictions could be related to this aspect of model building, or “storytelling” through processes of manipulation, as we will learn through pandemic modeling.

Following Espeland’s and Steven’s account on quantification practices, modeling as a measuring practice aims at controlling and predicting risks through quantification. The modeled encounters with risk, after all, are encounters to minimize the risk, to predict, and to prepare in front of the uncertain course of events. In broader terms, both types of prediction, explanation-based and scenario building, are *technologies of governance* that allow different interest groups to act at a distance (cf. Miller and Rose 2008). In explanation-based predictions, the underlying uncertainties are smaller, perhaps more manageable, whereas in scenario-building predictions the distance between what is known and what remains unknown is greater. Scenario-building predictions share some similarities with audit process, as discussed in Power (1997, p. 40):

- ▶ The audit process shrouds itself in a network of procedural routines and chains of unverified assurance, which express certain rituals of evidence gathering, but which leave the basic epistemic problem intact.

But are these similarities actually showing us what may result from overreliance on regulatory processes of governance? As we will learn through the case study of scenario-building predictions, their capacity to explain the phenomenon may manifest as a limitation or restriction, and yet they operate as useful tools to shed light on unknown future state of an anticipated public health risk. In the following, I will study in detail the modes of prediction provided by models. Through the analysis, I will show how useful models are in encountering and governing public health risks.

## Predicting Infectious Risks Through Modeling

In public health decision making, predicting is one of the key motivations to develop modeling techniques. What kinds of model-based predictions we are able to identify in infectious disease studies? Two cases analyzed in this chapter allow us to compare different types of model-based predictions (cf. Mansnerus 2011b). First, as an example of predictions that facilitate the renewal of vaccination strategies, a case of population-level transmission models of *Haemophilus influenzae* type b bacteria is analyzed. This case introduces us to *explanation-based* predictions that produce “what would happen if” scenarios. These scenarios derive their predictive capabilities from the available datasets and reach out to short-term predictions

beneficial to predict outbreaks within a particular area. So, the development of preventive measures in public health can be informed by *explanation-based* predictions.

Secondly, by analyzing a microsimulation model on mitigation strategies for a pandemic influenza, we will learn about *scenario-building* predictions. Typical for these predictions is that the data utilized in them are derived from past pandemics. Hence, these predictions are not capable to explain a possible future pandemic, but to produce reliable scenarios of its potential development, and thus facilitate the distribution of protective measures. So, in order to assess reliability and usability of model-based predictions, it is beneficial to increase transparency of evidence throughout the production and utilization process. This allows the different groups, who are involved in the decision-making processes, to evaluate the predictive scenarios and make well-informed decisions.

However, within infectious disease studies, one of the major public health concerns is the limited capability to predict emergence of outbreaks and people's behavior in such an event. Outbreaks could be regarded either as "small," when they occur, say in closed populations, such as army units, or "large," such as the anticipated pandemic outbreak. Small outbreaks, say transmission of bacterial meningitis, caused by Hib, in a military garrison may not receive broad media coverage, but are nevertheless important for the core tasks of public health officials. After all, it presents a life-threatening risk. To protect public health asks to be prepared for or capable of controlling and managing these outbreaks. Dynamic transmission models provide a rather flexible tool in order to do that – they form a ground to address anticipatory "what would happen if"-type questions. Larger, unexpected outbreaks that are capable to cause wider devastation gain easily significant attention. Preparedness plans are conducted both on national and international level. Large-scale simulation models that utilize data from past pandemics, on travel patterns and population density, produce a part of the scientific evidence base. One example of these models focuses on mitigation strategies and provides estimates of their effectiveness. So, these two cases analyzed in this study inform us of the two distinct modes of prediction represented in the models.

## Explanation-Based Predictions

---

Infections that affect mainly children's health are a mundane public health concern. One of the main threats is considered to be bacterial meningitis, because of its life-threatening nature. However, most of these infections are vaccine preventable, as in our example case, *Haemophilus influenzae* type b bacterial transmission. The main effort remains to reduce the risk of these severe disease forms in a population. So, the need to predict potential public health risks is answered by developing sophisticated transmission models. Evidence of potential outbreaks, indirect effects of vaccinations, and estimates of herd immunity are assessed by models. What kinds of predictions are useful to form the evidence base for vaccine-preventable infections? Amy Dahan Dalmenico (2007) argues that there is a continuous tension between the explanatory and predictive functions of models. According to her, this tension is seen as a source of conflict and compromise:

- ▶ Modeling practices [...] should they be first and foremost predictive and operational or cognitive and explanatory. Tension between explanatory and predictive capacities, between understanding and forecasting is a source of conflict and compromise in modeling (Dahan Dalmenico 2007, p. 126).

From a philosophical point of view, the distinction between explanations and predictions is considered separate or even in a conflict with each other as Dahan Dalmenico suggests. However, my analysis of the short-term, *explanation-based* predictions in the case of Hib transmission models, will argue that the tension could be set aside. Model-based predictions can be grounded on explanatory mechanisms, as the case of Hib transmission models, or they can provide desirable qualitative tools in a form of *scenario-building* prediction, as we will see. This is an interesting outcome, and useful when we are looking at how predictions help in public health risk assessment. The main benefit from models is that they allow us to “do things with numbers,” to build the platform upon which one can develop understanding of the infectious risk itself and experiment with the various mitigation strategies. Through these quantifiable tools, the evidence base for risk governance is widened. In the following, I will present a detailed case study how *explanation-based predictions* work in the case of modeling Hib transmission.

## Case of an Integrated Simulation Model of *Haemophilus influenzae* Type B (Hib) Transmission

Hib colonizes the human nasopharynx and is transmitted in droplets of saliva. The public health risk is related to its severe disease forms (Ladhani et al. 2009). Hib is capable of causing severe and often life-threatening diseases, such as meningitis and pneumonia in young children (an estimated three million cases of serious illness and 400,000 deaths each year in children under 5 years of age worldwide). A part of the incentive to produce model-based predictions lies in the cost of vaccines. Hib vaccine is not yet a part of national vaccination strategies in the developing countries, mainly in Africa and Asia. Polysaccharide vaccines were on market in the 1970s and conjugates in the 1980s. The main difference is that the polysaccharides protect against the disease forms, whereas the conjugates are capable of reducing the carriage of the bacteria and hence have effect on population level circulation of the bacteria. If considered from the economic point of view, polysaccharide vaccines are older and somewhat cheaper to produce, and the conjugates are more expensive. As clarified by Hib Initiative, Hib infections are difficult to treat in the developing countries, due to the lack of access to antibiotics, which are proven to be effective when treating the severe disease forms ([www.hibaction.org](http://www.hibaction.org), accessed 25.3.2009). Because of this, the Hib Initiative presents an estimate that 20% of children in developing countries with meningitis caused by Hib will die and 15–20% of children suffering from it will develop lifelong disabilities. As an epidemiologist from the Helsinki modeling group argues:

- ▶ WHO and GAVI (the Global Alliance of Vaccinations and Inoculations) advocate Hib conjugate vaccines, the major question remains whether universal vaccination will be at all feasible in the poorest economies. Will it be cost-effective, and will it be an appropriate use of resources among other possible health interventions? Schedules optimizing the age of vaccination and the number of doses are crucial for the acceptance of the expensive vaccines (Leino 2003).

These general concerns are translated into an integrated simulation model in order to produce qualitative, anticipatory predictions of the potential vaccination effects on the population level. The translation process meant that the modelers needed to study particular

mechanisms that were responsible for the behavior of the bacteria. In order to address these mechanisms in the integrated model, they studied them separately.

The global concern to implement conjugate vaccines is based on data from the UK and Finland. Both countries tell their own “success stories” that support the initiative to include Hib conjugate vaccines in the vaccination programs.

## “What Would Happen If” Questions as a Key to Explanation-Based Predictions

---

Seeking answers to “why”-questions means *explaining* a particular phenomenon, say a cause of an infection. When “why”-questions are addressed in models, they search for a particular mechanism that is responsible for the phenomenon. In other words, models capture epidemiological mechanisms and extrapolate explanations on the basis of that. But what are mechanisms and how are they addressed in models?

In order to develop the notion of explanation-based predictions as anticipatory techniques to address public health risks, I will discuss how the mechanism of natural immunity was expressed in a population-simulation model in order to gain short-term predictions to assess the efficacy of Hib-vaccines. So, the short-term predictions that answer “what would happen if” questions, even though studied in the Finnish context provide a potentially broader application context when extended or applied to address the benefits of implementing Hib vaccines in the developing countries.

In general terms, *explanation-based* predictions are predictions that explain the causal mechanism(s) responsible for a particular phenomenon and extrapolate on the basis of that short-term predictions, i.e., answers to “what would happen if”-type questions. In order to unpack this, I will elaborate the role of mechanisms and their relation to explanation-based predictions. *Mechanisms* form the basis or *anchor* the explanations to the available datasets, the epidemiological ground of the phenomena. Bechtel and Abrahamsen (2005, p. 423) define a mechanism as follows:

- ▶ A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena.

This definition clearly underlines that mechanism is involved with orchestrated functioning, which I interpret as being capable of bringing together specific properties, parts or operations of the phenomena. Mechanisms are responsible for a phenomenon, mobilizing its cause, occurrence, or development. In this sense, mechanisms contain the generalizable properties of the phenomena.

Disease transmission is a multiplex phenomenon, which is dependent, for example, on the frequency of contacts within a population group, infectivity of the pathogen, and the existing immunity within the population. These aspects of the transmission were taken into account, when a mechanism was explained in a model. In other words, studying research questions in the family of Hib-models helped clarifying the disease transmission mechanism and uncovered the connection between a mechanism and the research questions addressed in modeling. These models were built during 1994–2003 within the research collaboration between the National Public Health Institute and the University of Helsinki.

Let us study more closely how *explanation-based* predictions were established in a population-simulation model.

The leading question motivating the building of the population-simulation model was: What would happen if a 5-year-old child  $x$  acquires a Hib infection and how likely she is to infect the members of her family? This question is by its nature a “what would happen if” question that has a predictive emphasis. To see how this question was manipulated in the model, we need to unpack the structure of the model itself. The population-simulation model, published in 2004, has a three-part structure: a demographic model (covering the age-structure of a Finnish population), a Hib-transmission model (including the contact-site structure), and an immunity model (including the immunization program and its effects). Yet, this simulation model resulted after a 10-year period of modeling work, which was dominated by integrating practices that brought together the three parts, built earlier in the project (see Mattila 2006c). So, all three parts, especially the transmission model and immunity model, were partially studied prior to the accomplishment of the population-simulation model (2004) in terms of following questions (the year after the question refers to the published model):

- How long does the immunity [against Hib] persist? (1999)
- How do we estimate the interaction between the force of infection and the duration of immunity? (2000)
- What is the effect of vaccinations? (2001)

These questions address particular aspects that affect the transmission dynamics in a population: length of immunity, estimate related to the force of infection, and effect of vaccinations. In particular, two mechanisms were detected in these models: the mechanism of immunity and mechanism of transmission. Mechanism of immunity was defined as:

- ▶ Natural immunity is believed to depend on repeated exposure to Hib bacteria resulting in the production of functional antibody (Leino et al. 2000).

This mechanism is primarily about how to sustain natural immunity in a population. In the simulation model, it was used for explaining what would happen to the natural immunity, when vaccinations were introduced on a population level. This was an important aspect, since the epidemiological studies of the chosen vaccine confirmed that the vaccine itself is capable of reducing carriage. The reduction of carriage in a population could potentially lead to the waning of the natural immunity that had protective impact on a population level. In other words, *herd immunity* (the population level protection against an infection) could be affected (cf. Fine 1993). This indirect effect was documented in the model studying the dynamics of natural immunity. This mechanism and its numerical estimates, which were defined in terms of Hib antibody dynamics, show the descending trend in serum antibody concentration. Later, this mechanism was integrated in the population-simulation model, in particular into its immunity model part. Hence, the mechanism of natural immunity, when manipulated in the simulation model, showed that if the bacterial circulation is diminished, the natural immunity is likely to weaken and a potential increase in the risk of serious infections may affect those who are not vaccinated.

Explanation-based predictions hence allow us to both explain the phenomenon of interest and predict in a short-term its development, i.e., the course of Hib transmission in a population and the underlying epidemiological mechanisms that maintain circulation of the bacteria. An interesting parallel can be drawn to den Butter and Morgan (2000, p. 296):

- ▶ More general empirical models provide a consistent and quantitative indication of the net outcome of the various principle mechanisms thought to be at work based on the particular case (not stylized facts) and which might be affected by the policies proposed.

As den Butter and Morgan show, empirical economic models are linked with mechanisms as well. These models provide a base to work on a particular case and examine what kinds of effects suggested policies have. In a similar way, explanation-based predictions in public health policies allow estimations of risks by showing the short-term development of the infections, explicating the optimal immunity levels within the community, and sometimes even providing unexpected results of the optimal vaccination coverage. This was discussed in a lecture by Auranen and Leino (Lecture given at the London School of Economics, Workshop organised by the Economic History Department, March 2008), when they showed that Hib conjugate vaccine minimizes the carriage of the bacteria and allows optimization of vaccine coverage to be as low as 10%.

## Building Pandemic Scenarios

---

*Explanation-based predictions*, as discussed above, provide the ideal ground for short-term anticipation of public health risks, or low-impact, high-frequency events, as referred in the risk literature (cf. Hutter and Power 2005). However, most of the media attention is given to high-impact, low-frequency events, which in the public health context are pandemics. How do we respond to these events? Following International Health Regulations (IHR were revised by the WHO in 2005), each country is responsible for notifying WHO of “any events that may constitute a public health emergency of international concern.” In a way, these internationally coordinated activities are an early warning, but they may not be able to anticipate or predict the occurrence of a pandemic. According to WHO, we are currently living in a pandemic period, which means that preparedness plans are in use on national and international level and predictive models are tinkered with new daily estimates of the course of the pandemic.

How do *scenario-building* predictions form a part of the scientific evidence base for decision-making? *Scenario-building* predictions are predictions that “sketch, outline or describe an imagined situation or sequence of events, and outline any possible sequence of future events” (OED). In other words, *scenario-building* predictions are primarily tools to *produce qualitative scenarios* based on the available, past data, and as such they provide model-based encounters with future risks. These scenarios are not necessarily grounded directly on data of the future event (which does not exist), but build upon available sources of past data in order to anticipate the “unknown,” the risk.

## Predicting the Pandemic

---

Humankind has faced cycles of pandemics, one of the most famous being 1918 Spanish flu that killed, according to older estimates approximately 50 million people worldwide. The pandemic spread all around the world and lasted about 2 years (1918–1920). Its oddities were that it infected and killed young and healthy, and it spread during the spring months. The most recent cycle of a pandemic began in the end of April 2009, when human cases of a novel influenza type A virus were confirmed. These cases were identified in the USA and in Mexico. The virus, according to epidemiological evidence, had been circulating in Mexico since February 2009

and may have emerged already earlier that year. It was also confirmed that the new human strain was identical to a strain of virus that had been circulating in pigs in North America. Flu survey reports that the A(H1N1) strain has a complicated history: “some of its genes moved to birds to pigs in 1918, other genes from birds to pigs at the end of the 20th century, some got into pigs in the 1960s having first passed through humans.”

The strain spread rapidly, the first infections happened through contacts with those who were or traveled from Mexico. WHO reacted to the public health emergency by raising the Pandemic Alert Level from 4 to 5 (sustained community outbreaks in a limited number of countries) at the end of April. On June 11, 2009, WHO declared a pandemic and raised the Alert Level to phase 6, which means wide geographical spread, but does not indicate the severity of the infection.

According to ECDC Situation Report (27.7.2009), within the EU/EFTA countries, there are 20,512 confirmed cases and 35 deaths among those cases. Outside EU/EFTA countries, the corresponding numbers are 139,526 confirmed cases and 956 deaths. So far, critical voices have questioned the rationale of the pandemic alertness, since the cases seem to be somewhat mild and responding to the antiviral treatments. The major concern, however, was that there is very little natural prior immunity to the new strain and the infection it causes. This was already seen in the fact that the main group of infected is children. Due to the uncertainty of how serious the new type of virus was, the information campaigns for increased hygiene, advice for general audience and risks groups were available. In July 2009, vaccine production was underway, first vaccines were available for risk groups in September 2009, and, for example, the UK bought 90 million doses of vaccine in order to vaccinate the whole of population.

Uncertainties of the severity and spread of a pandemic raise questions of how to develop mitigation strategies to protect populations. Simulation models provide a way to predict the possible future course of the pandemic flu and hence function as a tool for planning and testing intervention strategies. When the simulation techniques are used in the preparedness planning, the data are grounded on observations from the past pandemics (1918 and 1957). These predictive simulation models allow studying various mitigation strategies.

## What Kinds of Models Are Used as Scientific Evidence Base for Preparedness Planning?

One of the major public health concerns in infectious disease studies is the limited capability to predict emergence of outbreaks and people’s behavior in such an event. To mitigate this problem, several studies have developed large-scale simulation models that utilize data from past pandemics, on travel patterns and population density. In the following, I will focus on one rather recent pandemic flu model and discuss its predictive capabilities. The model in question is an individual-based simulation model of pandemic influenza transmission for Great Britain and the United States (Ferguson et al. 2006). It represents transmission in households, schools and workplaces, and the wider community. The main aim of the model is to study strategies for mitigation of influenza pandemic. Mitigation means all actions that aim at reducing the impact of a pandemic (Nicoll and Coulombier 2009). I will focus on two model-based assumptions that affect the transmission: estimate for the reproductive rate and behavior. On the basis of a closer analysis of these assumptions, I will discuss the nature of *scenario-building* predictions and especially reflect on the suggested policy outcomes of this model.

## Fingerprint of the Pathogen, and of the Population

---

Transmission is quantified in epidemiological models as a basic reproductive rate, which is the rate that is used for estimating the spread of infection in a susceptible population. It is defined as  $R_0$ , which is the average number of individuals directly infected by an infectious case during her entire infectious period, when she enters a totally susceptible population. In infections that are transmitted from person-to-person, the potential of the spread is called the reproductive rate that depends on the risk of transmission in a contact and also on how common the contacts are. The reproductive rate is determined by the following four factors (Giesecke 2002):

- The probability of transmission in a contact between an infected individual and a susceptible one
- The frequency of contacts in the population
- How long an infected person is infectious
- The proportion of the population that is already immune

All these characteristics can be expressed in mathematical equations to provide numerical estimates of the transmission dynamics in a population. This rate is usually determined by empirical data, i.e., by deriving the estimate from previous epidemiological studies. However, it is a rate that carries a “fingerprint” of the pathogen. By this I mean that the reproductive rate is sensitive to particular strain of the pathogen in question. This sensitivity brings in a question of uncertainty in the model-based predictions. What if the strain is not so virulent? Alternatives are taken into account by modeling different possible scenarios based on different approximates of the reproductive rate. But what do the models do to the reproductive rate? In pandemic flu modeling, a future strain is unknown and therefore the models actually use data from the past strains. This relies, of course, on the assumption that the future pandemic is as virulent and contagious as the past one. If we look more closely to the reproductive rate and its variation, we can see how it manifests itself as a fingerprint of the pathogen. This idea means that population density affects the estimate since  $R_0$  tends to be higher in crowded populations. Nicoll and Coulombier (2009, Table 4) provide following estimates for  $R_0$ :

- In seasonal influenza:  $R$  around 1.1–1.2
- In pandemic influenza:  $R = 1.5–2.5$
- In current pandemic (H1N1):  $R = 1.5–2$
- In measles:  $R_0 > 10$

The variance in  $R_0$  leaves uncertainty into the predictions. This uncertainty is decreased once the pandemic begins to spread, and the pathogen is isolated and its virulence within a population (e.g., who are encountering the infection) is known.

## Behavioral Assumptions and Their Alternatives

---

The simulation model studying strategies for mitigating influenza pandemic makes assumptions concerning the effectiveness of behavioral interventions. These are movement restrictions, travel restrictions, quarantine, and school closure. The question is: What kinds of behavioral assumptions are made in order to predict the spread and transmission of the outbreak?

In Ferguson et al. (2006), a rather clear behavioral assumption is claimed when reporting the model design:

- We do not assume any spontaneous change in behavior of uninfected individuals as the pandemic progresses, but note that behavioral changes that increased social distance together with some school and workplace closure occurred in past pandemics.

Furthermore, the underlying assumption is to consider that individuals will behave according to the guidelines, rules, and restrictions given by the health authorities. In a way, the effectiveness of behavioral restrictions is based on the assumption of rational agents. But how reliable this assumption is? In a recent discussion on the novel ways to study real-world epidemics, Eric Lofgren and Nina Fefferman (2007) suggest that virtual game worlds might provide a different perspective. According to their analysis of an outbreak in an Internet playground, World of Warcraft, they observed that individuals did not follow the rules of movement restrictions and some voluntarily spread the disease. The question is: If the scientific simulation models are used for preparedness planning, how do we find reliable assumptions concerning the behavior, which is, after all a key to prevent the spread of pandemics?

## Scenario-Building Predictions

---

What is, then, the policy outcome of the model? What kind of scenarios the model suggested? Both epidemiological and behavioral assumptions have their limitations. On the epidemiological level, the assumptions represent the *fingerprint* of the pathogen, hence leaving some level of uncertainty when drawn to the predictive scenarios. On the behavioral level, the assumption that individuals' behavior remains unchanged during the pandemic period opens the questions of credibility of these scenarios. Yet, it was clearly stated that the models allowed to explore “number of scenarios” regarding the transmissibility of the pathogen, movement, and travel restrictions. One could easily think that if *scenario-building* predictions are relying on particularly uncertain assumptions, they are mere fantasies, no better than “fortune-telling.” However, this is not the case. As documented already with the Helsinki models on Hib, models provide a useful “playground,” a platform to examine and explore particular features of the infection and its transmission (cf. Keating and Cambrosio 2000, 2003; Mattila 2006c).

*Scenarios* which allow us to “access the inaccessible” provide qualitative tools and produce evidence of the unpredicted for decision making. The challenge remains how to communicate this particular mode of evidence – its changing and mutable nature (cf. Mansnerus 2011a). As Ferguson et al. (2006, p. 451) state: “The transmissibility of a future pandemic virus is uncertain, so we explored a number of *scenarios* here.” They argue that these scenarios depend on “model validation and parameter estimation,” which should be given a priority in future research. Transmissibility, which is based on the estimate of the reproductive rate, is considered to be on the level of 1918, if it actually follows the levels seen in 1968 or 1957 pandemics, “global spread will be slower and all the non-travel-related control policies examined here will have substantially greater impact.” Ferguson et al. emphasize the importance to collect the “most detailed data on the clinical and epidemiological characteristics of a new virus.” In other words, he is calling for research that allows us to base the scenario-building into a detailed understanding of the explanatory mechanisms of phenomena. The “fingerprint” of the pathogen is important, as pandemic simulations show.

What kind of scenario was built around the behavioral assumption? Interestingly, the outcome of the simulation model suggests that travel restrictions, which include both border controls and within-country restrictions, “achieve little” in delaying the peak of the epidemic. This was taken into account when WHO gave recommendations and guidance on traveling during the current A(H1N1)v pandemic:

- ▶ Scientific research based on mathematical modeling shows that restricting travel would be of limited or no benefit in stopping the spread of disease (7.5.2009, WHO, GAR, Travel: Is it safe to travel?).

The social functions of simulation models are also worth emphasizing. Scenario-building helps allocate resources, agree on, for example, preordering and manufacturing the vaccines, and stocking the antiviral. As we observed in the two examples, the scientific models have “uncertainty” built-in: the assumptions made on the basis of past facts may not provide accurate predictions of the scale of the outbreak. Nor are they capable of capturing the changing behavioral patterns of individuals. Testing out both assumptions and exploring them as part of various scenarios was “doable” only by modeling. This is an indication of the usefulness of *scenario-building* predictions; they are qualitative tools that “fill the gaps” in existing knowledge, allow reasoning to touch upon the “known unknowns” and perhaps “unknown unknowns.” A good example of “unknown unknowns” was the origin of outbreak of A(H1N1) pandemic in 2009. The main focus was on avian (H5N1) influenza that is currently circulating in South-East Asia. However, the pandemic emerged from pig farming industry in Mexico (cf. Mansnerus 2010).

## Beneficial Encounters with Infectious Risks

---

How modeling provides useful tools for risk assessment? As we learned through the case studies, predictive capacities of models encompass the ambition to “access the inaccessible,” as Oreskes (2007) shows in her analysis of scale models in geology. She refers to models “whose predictions are temporally or physically in accessible.” But this ambition is not merely epistemic. Models, either physical or numerical, seemed not to adjust to the changes in epistemic values shared in scientific communities, but also reflect aspirations of scientific patrons, as Oreskes discusses. Models seem to do more than epistemological work: in the attempts to predict the future, models generate predictions to inform policy decisions. This is what Oreskes argues to be the primarily social role of predictions. In a way, *scenario-building* capacities of models express the social role by providing “access to the inaccessible,” even though scenarios may not satisfy the epistemic quest of explaining the viral mechanisms of a pandemic.

In scenario-building predictions, the epistemic, for example, the precise rate of transmissibility plays a secondary role. The main importance is to explore and evaluate various outcomes. On the contrary, explanation-based predictions, when they successfully encompass epidemiological mechanisms, accommodate both the epistemic and social functions.

What kinds of modeled encounters with public health risks do the two types of predictions provide? How reliable are they? The analysis supports Boumans’s (2004) notion of instrumental reliability, which incorporates both the “instrument” and expertise required. In other words, reliable predictions, in both cases, result as the quality of the model and the expertise of the modelers. This means that calibration of the model is not indifferent to the other factors,

expertise of building the model and practice of using the model when addressing instrumental reliability. Oreskes and Belitz (2001) make a similar point by arguing that all models are approximations, and they suggest that it is more useful to “think of models as tools to be modified in response to knowledge gained through continued observation of the natural systems being represented.” In other words, when estimating public health risks through model-based predictions, instrumental reliability refers to the fact that these predictions are not valid descriptions of reality, but best available approximations of the risk. They are not static either, but as more data are cumulated during the outbreak, they gain greater accuracy. Both types of predictions show that modeled encounters with public health risks depend on the complex chain of interactions between experts and technologies, and between users and producers of these predictions.

*Explanation-based* predictions that are utilized in assessing low-impact, high-frequency infectious risks function in two ways. First, they explain the phenomenon by allowing researchers or policy-makers manipulate the model by questions. As we learned, the broader policy-driven questions are translated during the modeling process into smaller and more targeted questions that reveal the details of the transmission dynamics in a population and explain how the disease mechanism affects the possible infectious outbreak. Second, the explanation-based predictions are able to address “what would happen if”-type questions that arise when infectious outbreaks are encountered within a small group of population, such as a nursery group or military garrison. The prediction as an answer to “what would happen if” question is beneficial for assessing risk and further mitigation strategies, such as containment of the outbreak.

*Scenario-building* predictions should not be regarded as “nonsense” despite my choice to refer to the modeling platform that gives rise to them as a “playground.” As Oreskes pointed out, they function as “access to the inaccessible,” in that role they allow risk assessment to stretch itself beyond the “accessible”: Beyond the available data from surveillance or monitoring processes by simulating the outbreak on the basis of data from previous pandemics, or beyond the actual situation, i.e., ongoing outbreak by simulating variations of the spread of the infection and the effectiveness of mitigating strategies already during the pre-pandemic phase for preparedness planning. It seems that both these predictions provide beneficial tools to encounter public health risks from infections. However, their limitations are worth discussing in the context of analytics of governance.

## Further Research: Toward the Analytics of Risk

Modeled encounters with public health risks are proven highly beneficial, as we have learned. A challenge, however, remains to be tackled with, namely, the tension between population-level estimates of risk and individuals’ behavior. I will suggest this tension can be accommodated within the governmentality approach by showing how the analytics of risk benefits from the integration of technical and ethical rationalities along with the deepened understanding of risk rationality. This will be discussed as a direction for further research.

What are then the possible limitations of “modeled encounters” with risks? As the case of predictive scenarios of pandemics shows, availability of data may be limited or as in this case, nonexistent in regard to an actual outbreak when scenarios were built in pre-pandemic phase. Explanation-based predictions were also modeled on the basis of limited data, for in that case,

the data were collected for other purposes (as a part of pre-Hib vaccine studies) and therefore they did not accommodate all the relevant information for model parameterization. This meant that during the modeling, some parameter estimates were acquired on the basis of comparative datasets from collaborating research groups. Along with the limits of availability of data, computational capacity may present limits to modeled encounters with risks. The modelers may not be able to access the highest-level of computing power (such as supercomputers in national computing centers), which was the case with Hib-models. These technical limitations have an effect on the way in which models are built, how fluently a model-based prediction is gained, and how reliable the instrument, (i.e., the model) itself is. Limited access to high-level computational capacity restricted the number of simulation runs for the population-simulation model that estimated Hib transmission, for example. Although these restrictions may weaken the reliability of the model-based prediction, it is worth bearing in mind how Oreskes and Belitz (2001) described models as approximations.

Along with the technical limitations of modeled encounters with risks, there are social and epistemic limitations as well. As we learned through the analysis of the two types of models, the modeling process itself is not highly transparent. Specialized expertise is required to build the models, and even those who work with modelers may not be able to assess the choice of mathematical algorithms during the process. Interdisciplinary modeling teams develop a division of labor (see Mattila 2006a). This lack of transparency may be limiting when model-based predictions are communicated to audiences who have not been involved in the primary model building process or who are not familiar with modeling techniques. This is the point when models may turn into “truth-machines,” to gain their authority, as Hulme et al. (2009) suggests in the case of climate models. The assumptions made in the model may remain unknown due to the lack of communication of the modeling process and the choices made within it.

Furthermore, as I described that these models are typically *tailored* to address specific policy-driven questions, one could consider this characteristic a limitation. How applicable are the outcomes? If the simulation model particularly addresses a question like “what would happen if a child x in a day care unit y encounters a Hib infection?”, can the prediction be applied to estimate the risk of infection among adult men in military garrison? Or if the predictive scenario of a pandemic spread examines mitigation strategies, such as school closures or travel restrictions in a particular geographical location, can the estimates be extended to cover other areas as well? These questions address the inevitable limitations of modeled encounters with public health risks, which should not be read as a recommendation not to use modeling techniques or to advocate them. After all, model-based predictions, as these two cases show, are highly beneficial as a one source of evidence for the broader base of risk assessment. The limitations are discussed in order to balance the view.

As we learned, model-based predictions operate on population level. They provide information of risks that affect the whole population, hence being interested in the “welfare of the flock as a whole,” as Rose (2001) phrases Foucault’s terms. The “pastoral” attitude that is concerned of the welfare of the whole population, Rose continues, is a form of “collectivizing power.” This leads to a tension that arises when a public health risk manifests on a population level and appropriate health interventions are introduced, but at the same time, individuals consider their risk from a different angle and refuse or ignore to participate in the interventions. In other words, when an epidemic outbreak that causes a severe risk to the population (or to a part of it) happens, its further spread is prevented, for example, by vaccinations.

Yet, individuals may think that the side effects from the vaccination are more severe than the infection itself and refuse to follow the public health recommendations. But what lies behind this tension? I will reassess this tension from three perspectives: as a narrative, as a case of difference between individual's risk perception and that of a group, and as a challenge that needs a broader context to address it. The concept of narratives, as I already mentioned, is helpful when applied to modeling. Morgan talks about modeling as storytelling and I will follow the most recent work by Dry and Leach (2010) to discuss how to broaden out the modeled encounters with public health risks by acknowledging the narratives told through modeling. I will address the lack of focus on individuals' risk behavior, which was highlighted in the studies by Lofgren and Fefferman (2007). I will argue that a satisfying way to contextualize modeled encounters is by regarding them as *technical rationalities* within the governmentality approach to risk.

The fairly monolithic view of a population, as represented in the models, may lead to a biased interpretation of the model-based predictions or, more broadly, the model-outcomes and estimates. The population is seen as the "ultimate end of governance," as Rabinow and Rose (1994) claims. When governance seeks the form of modeling, we may use the metaphor of storytelling (Morgan 2001), which allows manipulation of the world through representing it in a model and addressing questions to it. But whose story is told and whose is ignored? Who remain silent? Dry and Leach (2010) raise this issue when they argue that narratives about infectious diseases are deeply rooted in questions of power and social justice. In order to address these questions, Dry and Leach suggest analyzing the different narratives that construct disease and epidemics. Narratives for them are not just stories, but stories with purposes and consequences. In a recent study on avian influenza surveillance, Scoones (2010) identifies three "outbreak narratives": A narrative that links veterinary risk with agriculture, a human public health narrative, and a narrative focusing on pandemic preparedness. His analysis shows that a single narrative is perhaps not enough to create the evidence base in order to understand the multiplicity of an infectious risk from pandemic. We could see the benefits of model-based predictions in a similar way. At best they give us a single narrative, and perhaps our task is to look for other complementary ones for well-grounded risk assessment.

One could take yet another step further and say that narratives, despite introducing more heterogeneity to the fairly fixed perspective on population, are still focused on groups rather than individuals. Neither *explanation-based* nor *scenario-building* predictions address individuals' perceptions of risks and the various factors that affect them. What is left aside in these modeled encounters with risks is the ways in which individuals perceive risk and how they behave. A typical bias in individuals' response to risk is known as *optimism bias*, which means that individuals underestimate risks to themselves (Costa-Font et al. 2009). Joffe (2003) argues that individuals construct risks through group attachment or on the basis of their experiences in groups. She continues that response to risk is therefore "a highly social, emotive and symbolic entity." Roeser's (2007) study on ethical intuitions about risks point to the same direction by acknowledging that individuals' intuitive risk judgments express ethical concerns that should be taken into account in methodologies for risk analysis or risk policy. None of these observations are accommodated on the population level analyses of public health risks.

How could we, then, satisfactorily accommodate the "unbearable tension" between individuals' perceptions of risks and the population-level assessment we gain through modeling? I will argue that we will benefit from a broader context to understand infectious risks in public health. By this broader context, I refer to the literature on governmentality that brings together

the technical rationalities of governance with ethical and epistemological aspects that are present in the process and manifest through the dynamics of power.

In Michel Foucault's work, the analytics of government covers three aspects that help contextualizing risk. These aspects focus on how we come to know about and act upon different conceptions of risk. How these different forms of risk rationality become a particular set of calculatory practices and technologies. How social and political identities emerge from these technologies (Dean 2010, p. 217). In this chapter, modeling techniques have been regarded as a form of technical rationality or *techne* in the governance of risk, which Dean defines as: "[...] a search for analytical clarity concerning the techniques and instruments of government, the arts, skills and means by which rule is accomplished" (Dean 1995, p. 560). As I have shown, model-based predictions are "instruments of government"; they are tools to anticipate risks, build predictive scenarios, and test mitigation strategies. At the same time, these techniques have their limitations. They easily enforce the purely probabilistic interpretation of risk, and the evidence produced by models assesses population-level risks but cannot include estimates for individual-level or address individuals' perception of risk simultaneously. This limitation can be addressed through the analytics of risk within the context of governance.

The analytics of risk is formed through four successive and overlapping stages, as Dean claims. In the beginning, one explores different forms of risk rationality, which Dean calls *episteme* of risk. Then, one seeks to find out how such conceptions are limited to particular technologies and practices that form the *techne* of risk. And finally, one studies how such technologies and practices give rise to new forms of social and political identity, and finally, how these identities are merged into political programs, which give them a particular *ethos* (Dean 2010, p. 217).

If we follow how Dean characterizes *episteme* within the analytics of government, we will notice a set of questions that are useful to map risk rationalities. "What forms of thought, knowledge, expertise, strategies, means of calculation, or rationality are employed in practices of governing?" (Dean 2010, 42). It seems to me that *episteme* and *techne* of risk rationality indicate a different direction or different dynamics of governing risks. Especially in the case of the pandemic, the risk considered to threaten the whole of population seemed to remain relatively small. Individuals made their own estimations for risk disregard to the recommendations or guidelines given by the public health officials. Those who were at risk did not consider the risk to be severe enough for them to follow the guidelines. As we can see, *episteme* and *techne* of risk rationality are pointing to the "care of oneself and of others," to the ethical dimension that is present when encountering public health risks. This forms the *ethos* of risk rationality or the social and political identities, which emerge out of *episteme* and *techne*, out of the rationalities of governance.

"Knowing an object is a process that shapes rationalities of governance by forming our understanding of how a risk of an infection is established and the ways in which all this was turned into a form of calculation," as Miller and Rose (2008, p. 30) define. For them, "knowing an object" involves "procedures of inscription," which are ways of collecting and presenting statistics, for example. It is not a process of speculative activity, but a way in which "governmentality" is made up. In my reading of governance of public health risks in the two cases, risk became a "knowable object." But the actors may not have reached a point what could be called "accountability of one's own actions." This is an important aspect of the *ethos* and its formation through rationalities of governance. Dean emphasizes that "if morality is understood as the attempt to make oneself accountable for one's own actions, or as a practice in

which human beings take their own conduct to be subject to self-regulation, then government is an intensely moral activity” (Dean 2010, p. 19). To recognize government as “an intensely moral activity” leads Dean to suggest that “techniques and rationalities of government needs to be complemented by a fuller clarification and elaboration around third axis, that for want of a better term, we might call ‘axis of self-formation’” (Dean 1995, p. 560). So, in order to complement risk rationality and to enhance successful governance of risk, I would suggest to include all “three axes of governmentality” into the process. This could lead to a balanced view which, according to Castell, seem to be threatened by modern ideologies.

- ▶ The modern ideologies of prevention are overarched by a grandiose technocratic rationalizing dream of absolute control of the accidental, understood as the irruption of the unpredictable. In the name of this myth of absolute eradication of risk, they construct a mass of new risks, which constitute so many new targets for preventive intervention (Castell 1991, p. 289).

Are the expiring stocks of pandemic vaccines a sign of “grandiose technocratic rationalizing of dream of absolute control?” This question asks for further research that potentially engages with critical assessment of risk governance and addresses the various tensions that may prevent good governance from reaching its purpose.

## Acknowledgments

---

I thank the British Academy for funding my research through Post Doctoral Fellowship at LSE Health, London School of Economics and Political Science. My research on risk and regulation was supported by an ESRC Post Doctoral Fellowship at the Centre for Analysis of Risk and Regulation (CARR), 2009. I was affiliated with Mellon Sawyer Seminar on Modelling Futures: Understanding Risk and Uncertainties at the Centre for Research in Arts, Social Sciences and Humanities, CRASSH, University of Cambridge, and a Visiting Fellow at Wolfson College, 2009–2010. I thank CRASSH for providing a dynamic research environment for me and Wolfson for hosting me.

## References

---

- Auranen K (1999) On Bayesian modelling of recurrent infections, Rolf Nevanlinna institute, faculty of science. University of Helsinki, Helsinki
- Auranen K (2000) Back-calculating the age-specificity of recurrent subclinical *Haemophilus influenzae* type b infection. Stat Med 19:281–296
- Auranen K, Ranta J, Takala A, Arjas E (1996) A statistical model of transmission of Hib bacteria in a family. Stat Med 15:2235–2252
- Auranen K, Eichner M, Käyhty H, Takala A, Arjas E (1999) A hierarchical Bayesian model to predict the duration of immunity to Hib. Biometrics 55(4):1306–1314
- Auranen K, Arjas E, Leino T, Takala A (2000) Transmission of pneumococcal carriage in families: a latent Markov process model for binary longitudinal data. J Am Stat Assoc 95(452):1044–1053
- Auranen K, Eichner M, Leino T, Takala A, Mäkelä PH, Takala T (2004) Modelling transmission, immunity and disease of *Haemophilus influenzae* type b in a structured population. Epidemiol Infect 132(5):947–957
- Bechtel W, Abrahamsen A (2005) Explanation: a mechanistic alternative. Stud Hist Philos Biol Biomed Sci 36:421–441
- Boumans M (1999) Built-in justification. In: Morgan M, Morrison M (eds) Models as mediators. Perspectives on natural and social sciences. Cambridge University Press, Cambridge
- Boumans M (2004) The reliability of an instrument. Soc Epistem 18(2–3):215–246

- Castell R (1991) From dangerousness to risk. In: Burchell G, Gordon C, Miller P (eds) *The Foucault effect studies in governmentality with two lectures and an interview with Michel Foucault*. The University of Chicago Press, Chicago, pp 281–298
- Costa-Font J, Mossialos E, Rudisill C (2009) Optimism and the perceptions of new risks. *J Risk Res* 12(1):27–41
- Dahan Dalmenico A (2007) Models and simulations in climate change: historical, epistemological, anthropological and political aspects. In: Creager A, Lunbeck E, Wise MN (eds) *Science without laws. Model systems, cases, exemplary narratives*. Duke University Press, Durham/London
- Dean M (1995) Governing the unemployed self in an active society. *Econ and Soc* 24(4):559–583
- Dean M (2010) *Governmentality: power and rule in modern society*, 2nd edn. Sage, London
- den Butter F, Morgan M (2000) Empirical models and policy-making. Routledge, London
- Dry S, Leach M (eds) (2010) *Epidemics, science, governance and social justice*. Earthscan, London
- Edwards P (1999) Global climate science, uncertainty and politics: data-laden models, model-filtered data. *Sci Cult* 8(4):437–472
- Espeland WN, Stevens ML (2008) A sociology of quantification. *Eur J Sociol* XLIX 3:401–436
- Evans R (2000) Economic models and economic policy: what economic forecasts can do for government. In: den Butter F, Morgan M (eds) *Empirical models and policy-making: interaction and institutions*. Routledge, New York
- Ewald F (1991) Insurance and risk. In: Burchell G, Gordon C, Miller P (eds) *The Foucault effect studies in governmentality with two lectures and an interview with Michel Foucault*. The University of Chicago Press, Chicago, pp 197–210
- Ferguson NM, Cummings DAT, Fraser C, Cajka J, Cooley P, Burke D (2006) Strategies for mitigating an influenza pandemic. *Nature* 442:448–452
- Fine P (1993) Herd immunity: history, theory, practice. *Epidemiol Rev* 15(2):265–302
- Foucault M (1978/1991) *Governmentality*. In: Burchell G, Gordon C, Miller P (eds) *The Foucault effect studies in governmentality with two lectures and an interview with Michel Foucault*. The University of Chicago Press, Chicago, pp 87–104
- Giesecke J (2002) *Modern infectious disease epidemiology*. Arnold, London
- Gramelsberger G (2010) Conceiving processes in atmospheric models – general equations, subscale parameterizations, and ‘superparameterizations’. *Stud Hist Philos Mod Phys* 41:233–241
- Habbema J, de Vlas S, Plaisier A, Oortmaassen G (1996) The microsimulation approach to epidemiologic modeling of helminthic infections, with special reference to schistosomiasis. *Am J Trop Med Hyg* 55(5):165–169
- Hamer WH (1906) *The millroy lectures on epidemic disease in England – the evidence of variability and of persistency of type*. Bedford Press, London
- Hillerbrand R (2010) On non-propositional aspects in modelling complex systems. *Anal Kritik* 32:107–120
- Hulme M, Pielke R Jr, Dessai S (2009) Keeping prediction in perspective. *Nature (Rep Clim Chang)* 3:126–127
- Hutter B, Power M (2005) *Organizational encounters with risk*. Cambridge University Press, Cambridge
- Joffe H (2003) Risk: from perception to social representation. *Br J Soc Psychol* 42:55–73
- Keating P, Cambrosio A (2000) Biomedical platforms. *Configurations* 8:337–387
- Keating P, Cambrosio A (2003) Biomedical platforms: realigning the normal and the pathological in late-twentieth century medicine. The MIT Press, Cambridge, MA
- Kermack WO, McKendrick AG (1927) A contribution to the mathematical theory of epidemics. *Proc R Soc Lond A Math Phys* 115(772):700–721
- Ladhan S, Neely F, Heath P, Nazareth B, Roberts R, Slack M, McVernon J, Ramsey M (2009) Recommendations for the prevention of secondary *Haemophilus influenzae* type b (Hib) disease. *J Infect* 58:3–14
- Leino T, Auranen K, Mäkelä PH, Takala A (2000) Dynamics of natural immunity caused by subclinical infections, case study on *Haemophilus influenzae* type b (Hib). *Epidemiol Infect* 125:583–591
- Leino T, Auranen K, Mäkelä PH, Käyhty H, Ramsey M, Slack M, Takala A (2002) *Haemophilus influenzae* type b and cross-reactive antigens in natural Hib infection dynamics; modelling in two populations. *Epidemiol Infect* 129:73–83
- Leino T (2003) Population immunity to *Haemophilus influenzae* type b - before and after conjugate vaccines. A23/2003. National Public Health Institute, Department of Vaccines. Helsinki, Finland
- Leino T, Takala T, Auranen K, Mäkelä PH, Takala A (2004) Indirect protection obtained by *Haemophilus influenzae* type b vaccination: analysis in a structured population model. *Epidemiol Infect* 132(5):959–966
- Lofgren ET, Fefferman N (2007) The untapped potential of virtual game worlds to shed light on real world epidemics. *Lancet* 7:625–629
- MacKenzie D (2005) Matematizing risk: models, arbitrage and crises. In: Hutter B, Power M (eds) *Organizational encounters with risk*. Cambridge University Press, Cambridge, pp 167–189
- Mäkelä PH, Käyhty H, Leino T, Auranen K, Peltola H, Lindholm N, Eskola J (2003) Long-term persistence of immunity after immunisation with *Haemophilus*

- influenzae type b conjugate vaccine. Vaccine 22:287–292
- Mansnerus E (2009a) Modelled encounters with public health risks: how do we predict the unpredictable? Published as a refereed discussion paper (No. 56) at the Centre for Analysis of Risk and Regulation, LSE, London
- Mansnerus E (2009b) The lives of facts in mathematical models: a story of population-level disease transmission of *Haemophilus influenzae* type b bacteria. BioSocieties 4(2/3):207–222
- Mansnerus E (2010) Ignorance and uncertainty in the life-cycles of evidence: the case of pandemic influenza preparedness planning. Published as a refereed discussion paper (No. 60) at the Centre for Analysis of Risk and Regulation, LSE, London
- Mansnerus E (2011a) Using models to keep us healthy: productive journeys of facts across public health networks. In: Howlett P, Morgan M (eds) How well Do ‘facts’ travel? Dissemination of reliable knowledge. Cambridge University Press, Cambridge, MA
- Mansnerus E (2011b) Explanatory and predictive functions of simulation modelling: case: *Haemophilus influenzae* type b dynamic transmission models. In: Gramelsberger G (ed) From science to computational sciences. Studies in the history of computing and its influence on today’s sciences. Diaphanes, Zuerich
- Mattila E (2006a) Interdisciplinarity in the making: modelling infectious diseases. Perspect Sci Hist Philos Sociol 13(4):531–553
- Mattila E (2006b) Struggle between specificity and generality: how do infectious disease models become a simulation platform. In: Kueppers G, Lenhard J, Shinn T (eds) Simulation: pragmatic constructions of reality, vol 25, Sociology of the sciences yearbook. Springer, Dordrecht, pp 125–138
- Mattila E (2006c) Questions to the artificial nature: A philosophical study of models in scientific practice. Thesis publication, University of Helsinki. Dark oy, Helsinki
- Miller P, Rose N (2008) Governing the present. Administering economic, social and personal life. Polity Press, Cambridge
- Morgan M (2001) Models, stories and the economic world. J Econ Methodol 8(3):361–384
- Morgan M (2002) Model experiments and models in experiments. In: Magnani L, Nersessian N (eds) Model-based reasoning: science, technology, values. Academic/Plenum, New York
- Morgan M, Morrison M (1999) Models as mediators. Perspectives on natural and social sciences. Cambridge University Press, Cambridge
- Nicoll A, Coulombier D (2009) Europe’s initial experience with pandemic (H1N1) 2009 – mitigation and delaying policies and practices. Euro Surveill 14(29): pii: 19279
- Oreskes N (2007) From scaling to simulation: changing meanings and ambitions of models in geology. In: Creager A, Lunbeck E, Wise MN (eds) Science without laws. Model systems, cases, exemplary narratives. Duke University Press, London
- Oreskes N, Belitz K (2001) Philosophical issues in model assessment. In: Anderson MG, Bates PD (eds) Model validation: perspectives in hydrological science. Wiley, London, pp 23–41
- Porter T (2000) Life insurance, medical testing, and the management of mortality. In: Daston L (ed) Biographies of scientific objects. The University of Chicago Press, Chicago, pp 226–246
- Power M (1997) The audit society. Rituals of verification. Oxford University Press, Oxford
- Rabinow P, Rose N (eds) (1994) The Essential Foucault. Selections from the essential works of Foucault 1954–1984. The New Press, New York/London
- Riesch H (2011) Levels of uncertainty. In: Roeser S et al (eds) Handbook of risk theory. Springer, London
- Roeser S (2007) Ethical intuitions about risks. Saf Sci Monit 3(11):1–30
- Rose N (2001) The politics of life itself. Theory Cult Soc 18(6):1–30
- Schllich T, Tröchler U (2006) The risks of medical innovation. Risk perception and assessment in historical context. Routledge, Abingdon
- Scoones I (2010) Fighting the Flu: risk, uncertainty and surveillance. In: Dry S, Leach M (eds) Epidemics, science, governance and social justice. Earthscan, London, pp 137–164
- Shackely S, Wynne B (1996) Representing uncertainty in global climate change science and policy: boundary-ordering devices and authority. Sci Technol Hum Value 21(3):275–302
- Smith RD (2006) Responding to global infectious disease outbreaks: lessons from SARS on the role of risk perception, communication and management. Soc Sci Med 63:3113–3123
- Taylor-Gooby P, Zinn J (eds) (2006) Risk in social science. Oxford University Press, Oxford
- van den Bogaard A (1999) Past measurements and future prediction. In: Morgan M, Morrison M (eds) Models as mediators: perspectives on natural and social science. Cambridge University Press, Cambridge, pp 282–326



# 10 Management of the Risks of Transport

John Adams

University College London, London, UK

<i>Introduction: “We Know What Works”</i> .....	240
<i>History: What Works?</i> .....	241
The 1962 Traffic Act .....	244
1967: The Breathalyzer .....	244
1973–1977: The Energy Crisis Speed Limits .....	245
1983: The Seatbelt Law .....	245
<i>Further Research – Explaining the Paradox</i> .....	248
Different Risk Managers .....	249
What Kills You Matters .....	252
Our Way of Life .....	254
What Sort of Risk? .....	255
Filters .....	257
Risk: An Interactive Phenomenon .....	258
<i>Conclusion</i> .....	260
A Concluding Speculation .....	262

**Abstract:** What does a transport safety regulator have in common with a shaman conducting a rain dance? They both have an inflated opinion of the effectiveness of their interventions in the functioning of the complex systems they purport to influence or control. There is however a significant difference. The clouds are indifferent to the antics of the shaman and his followers. But people react to the edicts of a regulator and frequently not in the way the regulator intends. There are two different kinds of managers involved in the management of transport risks: there are the “official,” institutional, risk managers who strive incessantly to make the systems for which they are responsible safer, and there are the billions of individual fallible human users of the systems, each balancing the rewards of risk against the potential accident risks associated with their behavior. Conventional road safety measures rest on a model of human behavior that assumes that road users are stupid, obedient automatons who are unresponsive to perceived changes in risk and who need protecting, by law, from their own and others’ stupidity. The idea of risk compensation underpins an alternative model of human behavior that road users are intelligent, vigilant, and responsive to evidence of safety and danger and, given the right signals and incentives, considerate.

## **Introduction: “We Know What Works”**

---

In March 2010, the United Nations proclaimed 2011–2020 the Decade of Action for Road Safety (UN announcement [2010](#)). It aspired to promote road safety everywhere, but especially in countries in the early stages of motorization with the highest accident rates.

Credit for this proclamation has been claimed by the Make Roads Safe campaign of the FIA Foundation ([Make Roads Safe 2010](#)). On the campaign’s Web site a “TAKE ACTION” tab offers a selection of Make-Roads-Safe T-shirts, banners, publications, and wristbands, but no suggestions for what might actually be done to make roads safer. Former UK defense secretary Lord Robertson, who is chairman of the campaign, is a bit more specific. He claims “We know what works: making vehicles safer and designing roads to be safe for all road users; tackling inappropriate speed and drink driving; promoting seat belt use and helmet wearing; improving driver training and police enforcement.”

The spearheading of the campaign by the FIA (Fédération Internationale de l’Automobile) the organization that glamorizes “inappropriate speed” through its promotion of Formula 1 racing, is an incongruity that seems thus far to have escaped comment elsewhere. At the time of writing (November 2010) the Make Roads Safe campaign is the most prominent feature on the FIA home page ([www.fia.com](http://www.fia.com)) – competing for attention with pictures of racing cars doing exciting things at inappropriate speeds.

At the launch ceremony for the campaign John Sammis, representing the United States, drew attention to the “6,000 of his fellow citizens killed and the more than half a million injured in 2009 due to distracted driving, particularly text messaging” (UN announcement [2010](#)). This contribution highlights a significant problem for those who claim to know what works.

In the United States, laws banning text messaging while driving are a matter of state jurisdiction; some states have passed laws others have not. This has created a natural experiment in which the accident experience of states with laws can be compared with the experience of those that have not. In 2010, the Highway Loss Data Institute published a report on the effect of the laws. It concluded:

- ▶ The results of this study seem clear. In none of the four states where texting bans could be studied was there a reduction in crashes. It is important to remember that the public safety issue in distracted driving is the crashes resulting from cell-phone conversations and texting, not the use of these devices, *per se*. If the goal of texting and cell phone bans is the reduction of crash risk, then the bans have so far been ineffective. Bans on handheld cell-phone use by drivers have had no effect on crashes (HLDI 2009), as measured by collision claim frequencies, and texting bans may actually have increased crashes (HLDI 2010).

The texting study concludes with a plausible speculation to explain the increase in crashes in *states that passed laws banning texting while driving*:

- ▶ This unexpected consequence of banning texting suggests that texting drivers have responded to the law, perhaps by attempting to avoid fines by hiding their phones from view. If this causes them to take their eyes off the road more than before the ban, then the bans may make texting more dangerous rather than eliminating it.

The perverse effect of texting bans created a difficulty for the US Department of Transportation Secretary Ray LaHood, a strong advocate of texting bans. He dealt with the difficulty by simply denouncing the study as “ridiculous” ([www.textkills.com/?p=1418](http://www.textkills.com/?p=1418)) and by issuing an angry, hand-waving dismissal of the method of assessment used by the HLDI, stating that the same method would have cast doubt on the efficacy of seatbelt and drink-drive legislation (USDOT 2010). As we shall see below this is a less than convincing argument. Like Lord Robertson, and numerous other road safety campaigners, LaHood knows what works and is exasperated by evidence that contradicts this “knowledge.”

The American experience with texting bans is but the most recent installment of a long-running saga. In 1985, the late Frank Haight, the long-term editor of *Accident Analysis and Prevention*, one of the most highly regarded scientific journals in the field, observed:

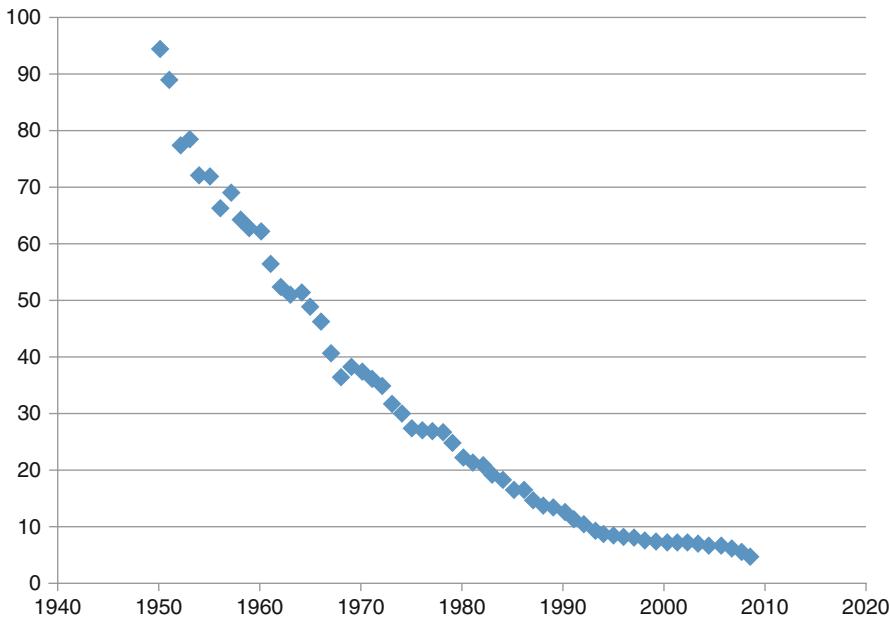
- ▶ One sees time and again large sums of money spent [on road safety] in industrialized countries, the effect of which is so difficult to detect that further sums must be spent in highly sophisticated evaluation techniques if one is to obtain even a clue as to the effectiveness of the intervention (Haight 1985).

## History: What Works?

Since the earliest days of mechanized transport there have been efforts to manage the risks that accompany it. In Britain the famous Red Flag Act (the Locomotive Act of 1865) required traction engines to be preceded by a man walking 60 yards ahead, at no more than 4 mph, carrying a red flag. This requirement was not repealed until 1896 – coincidentally the same year in which the first pedestrian was killed by a car in Britain.

Since then countless road safety measures have been implemented, in many jurisdictions: speed limits, accident black spot treatments, vehicle construction regulations, drink-drive laws, road signage, traffic lights, and seatbelt laws, to name but a few.

Concerns about the risks attached to transport have deep roots. The *Dublin Police Act* of 1842 created the offense of “driving furiously” – the same style of driving attributed to Jehu in the Old Testament (2 Kings, 20). As recently as November 2008, the Irish Law Commission

**Fig. 10.1**

Deaths per billion vehicle kilometers – GB

(Law Commission 2008) was consulting on whether this offense should be abolished. Whether it has now been abolished I am afraid that, at the time of writing, Google does not relate.

Intriguingly, despite decades if not millennia of interest in the problem of managing transport safety, there is remarkably little agreement about what works. Consider **Fig. 10.1**. Since 1950, road accident fatalities per kilometer traveled in Britain have dropped dramatically.

**Figure 10.2**, with the vertical axis transformed into logarithms, shows that the trend between 1950 and 2008 can be approximated by a straight line whose slope reveals that over this period deaths *per vehicle kilometer* decreased at a rate of 5.2% per year. The risk of death per kilometer traveled at the end of the period was 1/20th of the risk at the start of the period. Clearly something was working to make travel safer. But what?

The downward trend illustrated by **Fig. 10.2** does not mean that the number of road accident fatalities decreased every year. **Fig. 10.3** shows that in years when traffic grew a rate higher than 5.2% the number of fatalities tended to increase, and when it grew more slowly they tended to decrease.

The arrows on **Fig. 10.3** indicate (with the exception of the 1991 arrow) the introduction of significant road safety measures – government interventions intended to make the roads safer. Each should have produced, according to the prior claims of their promoters, a sharp downward step in the graph displayed in **Fig. 10.2**. But the steps are not there.

The first, the 1962 *Traffic Act*, imposed new speed limits, increased the maximum fines for speeding and careless driving by 150%, and introduced the “totting-up” procedure whereby

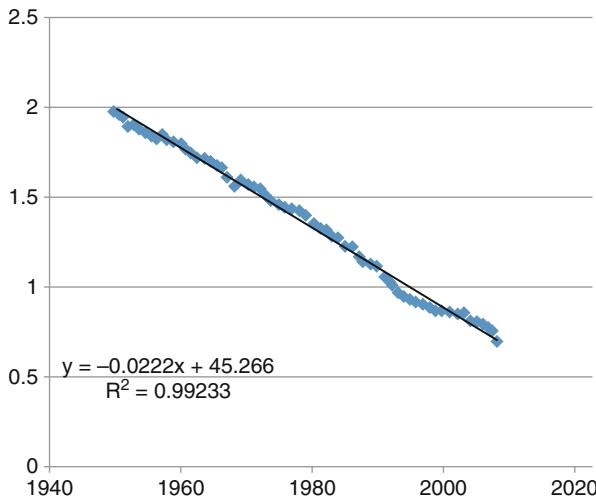


Fig. 10.2  
Log deaths per billion vehicle kilometers – GB

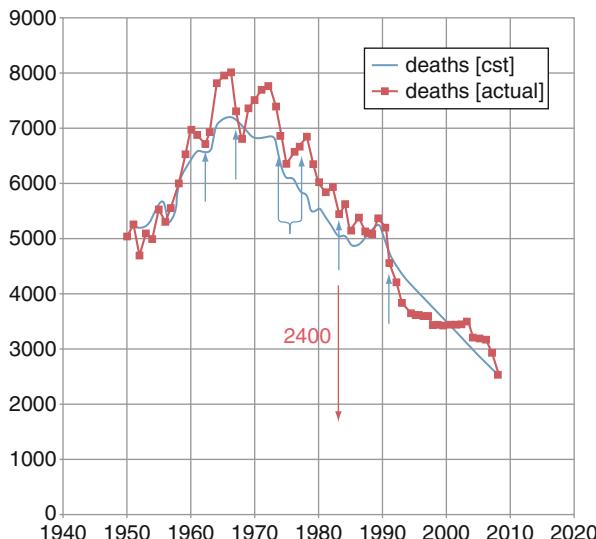


Fig. 10.3  
Road accident deaths in Great Britain: 1950–2008

drivers could be disqualified for three offenses. The second, the *Road Safety Act of 1967*, made it an offense to drive with over 80 mg of alcohol per 100 ml of blood. The third pair of arrows brackets the “energy crisis” speed limits. Between December 1973 and April 1977, various speed limits were imposed in response to the energy crisis and then repealed; they were introduced as

a fuel conservation measure, but were warmly welcomed by safety experts as a safety measure. The fourth, *the seat belt law* took effect in January 1983; it made the wearing of seatbelts in the front seats of cars and vans obligatory and was introduced with the claim that it would save 1,000 lives a year. The fifth arrow we shall come to shortly.

## The 1962 Traffic Act

---

The penalties provided by law for motoring offenses are intended to act as incentives to safer driving – disincentives to law-breaking being equated with disincentives to unsafe behavior. The objective of the new speed limits, larger maximum fines, and the totting-up procedure (whereby 12 penalty points led to disqualification) was to increase the severity of the punishment for the most persistent offenders. However, changing the law did not necessarily lead to a change in practice; although the maximum permitted fines for speeding and careless driving had been increased by 150%, the average fine handed out by the courts did not increase at all (Plowden 1971). Following the implementation of the measures contained in the 1962 Traffic Act, the number of road accident deaths, which had fallen over the previous 2 years, climbed more steeply than the trend identified in Fig. 10.3 until reaching a post-war peak in 1966. Perhaps the increase would have been even greater without the 1962 Act. Perhaps not.

## 1967: The Breathalyzer

---

The introduction of blood alcohol limits in October 1967 and a new method of testing coincided with a sharp drop in road accident fatalities. It appears likely that the new limits and the Breathalyzer deserve credit for a substantial part of this decrease. The number of over-the-limit dead drivers dropped from 25% to 15%. The number of deaths between 2200 and 0400 h (the period in which most drink-drive offenses are committed) dropped by 31%. However the effect was temporary. By 1969, the percentage of drivers killed in accidents while over the legal limit was back above its prelaw level. It is difficult to see a clear correlation between drinking and driving and total road accident deaths. By 1983, the number of over-the-limit dead drivers had risen to 31% while total road accident fatalities had dropped from 6,810 in 1968 to 5,445.

In 1983, *Accident Analysis and Prevention* devoted an entire issue to the problem of impaired driving. The guest editor summarized his long experience of drunken driving countermeasures in a despairing introduction:

- ▶ Once again, drinking and driving has come to the fore as a public concern. The beginning of every decade over the past 30 years has seen a surge of interest in, and concern over, drinking and driving. This concern has led to millions being spent throughout the world on countermeasures, with little measurable success in reducing the problem (Vinglis 1983).

It is frequently argued that the temporary success achieved by some drink-drive “blitzes” proves that the problem could be solved by some combination of more draconian penalties and more vigorous enforcement. Scandinavia, with its low permitted alcohol levels, rigorous enforcement, and draconian penalties for over-the-limit driving is frequently held up to the rest of the world as an exemplar. But Ross in a 1976 article entitled “The Scandinavian Myth”

(Ross 1976) cast doubt on this hypothesis. His interrupted time-series analyses revealed no effect of the Scandinavian drink-drive laws on the relevant accident statistics.

His analysis suggested that tough drink-drive legislation is only likely to work where it accords with prevailing public opinion. He noted the existence of a politically powerful temperance tradition in Scandinavia. Many people considered drinking and driving a serious offense (if not a sin) before it was officially designated as such by legislators. The absence of a detectable effect of Scandinavian drink-drive laws on accident statistics at the time the laws came into effect suggested, according to Ross, that the laws were symptomatic of a widespread concern about the problem, and that most people likely to obey such laws were already obeying them before they were passed. The laws, in effect, simply ratified established public opinion.

Where laws are passed that run ahead of public opinion there appears to be a conspiracy involving motorists, the police, judges, and juries to settle for a level of compliance and enforcement that accords with public opinion. In Britain after 1983, there was an impressive decrease in the number of dead drivers over the legal limit. The cause appears not to have been any specific intervention by the government, but a change in social attitudes.

### 1973–1977: The Energy Crisis Speed Limits

---

In December 1973, a blanket speed limit of 50 mph was applied to all roads in Britain not already subject to a lower limit. At the same time petrol prices were increased by 20%, followed by another increase of 20% in February 1974, and a further increase in April; between December 1973 and April 1974 petrol prices increased by about 57%. The motorway speed limit was restored to 70 mph at the end of March, and in May the 70 mph limit was restored to other all-purpose roads previously subject to that limit. In November 1974, the limit on some all-purpose roads was reduced to 50 mph and on others to 60 mph. Finally, in April 1977 Parliament agreed that the 50 and 60 mph limits on all-purpose roads should be raised again to 60 and 70 mph – in the face of protest and dire predictions by safety experts.

In 1974 and 1975, the total number of road deaths decreased. In 1976, 1977, and 1978 they increased. However the contribution of the different modes of travel to the changes in the total numbers of deaths in these years varied considerably. One response to the large increase in the price of petrol that accompanied the energy crisis was a large increase in the use of more energy efficient, but also more dangerous, motorcycles. Between 1975 and 1978, there was an increase of 465 in the total number of road deaths per year, but most of this increase (325) was accounted for by motorcyclists. In 1977, after the last of the energy crisis speed limits was repealed, the total number killed – excluding motorcyclists – decreased. After 1978, deaths for all modes, despite the dire predictions of road safety campaigners advocating lower speed limits, decreased markedly.

### 1983: The Seatbelt Law

---

The effect of the 1983 seatbelt law remains the subject of extraordinary myth making. On January 31, 2008, Britain's Department of Transport celebrated the 25th anniversary of the laws coming into effect with a press release in the name of the Road Safety Minister claiming:

- ▶ Twenty five years of seatbelt wearing laws have helped save 60,000 lives (Department for Transport 2008).

Others were quick to claim a share of the credit. The Web site of the Royal Society for the Prevention of Accidents claims:

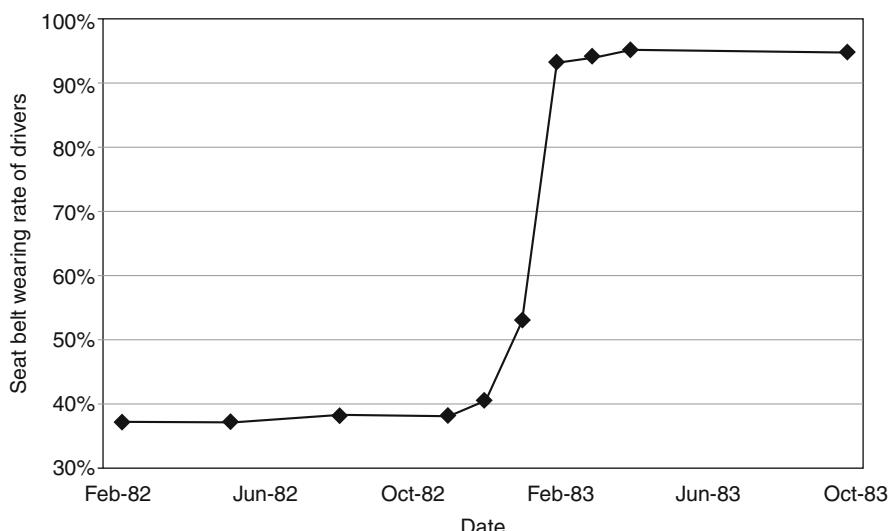
- ▶ 1982 – RoSPA's president, Lord Nugent, secured compulsory wearing of seatbelts with a late amendment to a Transport Bill. The law is estimated to have saved 60,000 lives to date (RoSPA 2010).

While the Parliamentary advisory Council on Transport Safety explained their role as follows:

- ▶ On the 31st January 2008, the 25th anniversary of the law change which made front seatbelt wearing compulsory was celebrated. PACTS itself was set up by Barry Sheerman MP as part of the fight to get mandatory seatbelt wearing turned into legislation. Eight years later it became compulsory for all backseat passengers to use seatbelts and it is estimated that since the introduction of the first law change in 1983, seatbelts have prevented 60,000 deaths and over 670,000 serious injuries (PACTS 2010).

The 60,000 claim has been endlessly recycled in the national and local press, radio, and television by the police and on Web sites including those of the National Health Service, insurance companies, law firms, and numerous others rather marginally connected to road safety concerns such as the Yorkshire Dales National Park. Such is the mesmerizing power of large numbers that the claim even escaped the usually sharp editorial eyes of a large team of highly experienced transport researchers who maintained in a report for the Department for Transport that “Over the past 25 years the compulsory wearing of seatbelts has been estimated to have saved at least 60,000 lives” (Erel Avineri et al. 2009).

Sixty thousand lives saved over a 25-year period averages 2,400 per year (shown on Fig. 10.3). The increase in wearing rates at the time the law came into effect was large and abrupt (see Fig. 10.4). The claimed effect of the law should have been evident in Fig. 10.3 as



**Fig. 10.4**  
Seatbelt-wearing rate of drivers

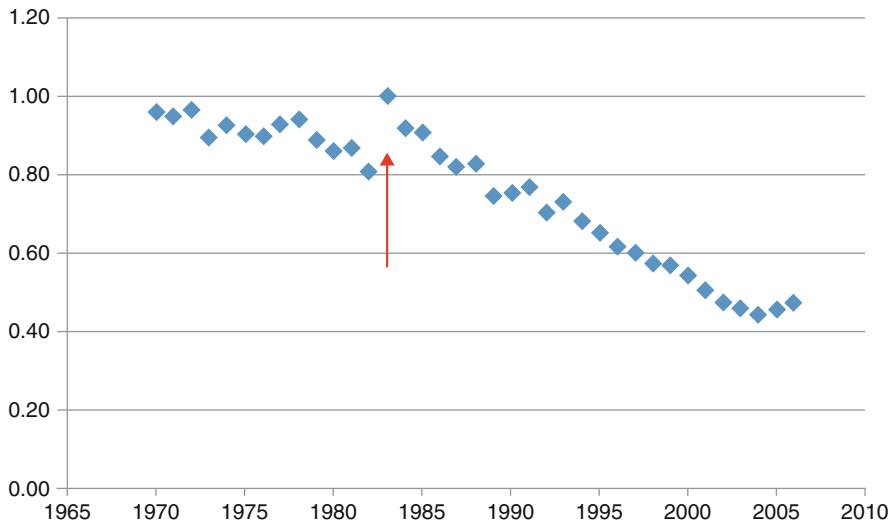


Fig. 10.5

Ratio of pedestrian and cyclist fatalities to car occupant fatalities 1970–2006

a sharp downward step in the established downward trend. Instead the trend leveled off, not resuming until after 1990.

There is however a sharp step effect to be seen in the road accident data. The ratio of pedestrian and cyclist fatalities to car occupant fatalities had been declining for many decades as the numbers traveling in cars increased and the amount of walking and cycling decreased. In 1935, the ratio was 6 to 1; by 1982 it was down to 0.8 to 1 (see Fig. 10.5). In 1983, it jumped 25% to 1.0 and it was another 7 years before it fell below 0.8. Consistent with the result in other jurisdictions with seat belt laws there was a shift in the burden of risk from those best protected in cars to more vulnerable road users on foot or bicycle (Adams 1995, 2010, Chap. 7).

In 1991, the total road accident fatalities decreased by 12.5%. This was the largest annual decrease since the war years when fuel shortages removed large numbers of vehicles from the roads. Frustratingly for road safety campaigners, it is not possible to attribute the decrease in 1991 to any of the safety measures introduced in that year. Indeed, 1991 was a quiet year on the road safety front in terms of the implementation of new safety measures. The following list presents the most significant new safety interventions in 1991 listed in *Road Accidents Great Britain 1991*, and the associated casualty effects, where available, from published sources.

#### Road safety measures implemented in 1991

- Twelve 20 mph zones introduced – *the decrease in casualties in built-up areas was less than the overall decrease*
- Thirty-one million pounds allocated for local safety schemes – *a sum equal to the value of 41 fatal accidents in a DoT cost-benefit analysis*
- Chevron markings tried out on the M1

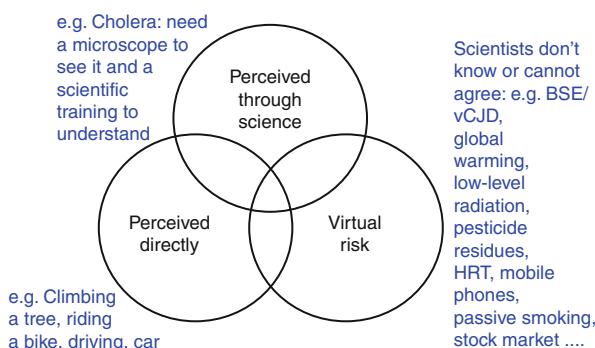
- Trials of nearside pedestrian signals at junctions
- Launch of “The Older Road User” campaign – *the decrease in casualties for those over age 65 was less than the overall decrease*
- Campaign to encourage wearing of cycle helmets by children – *decrease in cycling casualties ages 0–15 was less than overall decrease*  
Change in law requiring adults in rear seats to wear belts in cars where belts are fitted and available – *comparable statistics not available, but decrease in total rear seat casualties less than overall decrease*
- Campaign to encourage drivers to slow down in areas where children are likely to be about – *decrease in casualties suffered by pedestrians and cyclists age 0–15 was less than overall decrease*

A more plausible explanation for the record decrease in road deaths in 1991 than any actions on the part of the Department of Transport or other safety organizations is that the decrease coincided with the most severe recession since the war. There is clear evidence that road accident casualties go up and down with the economy (Adams 1985, Chap. 7).

## Further Research – Explaining the Paradox

► Figures 10.1–10.3 represent what for most transport risk managers is a paradox. They display an enormous decrease in the death rate per volume of traffic with no significant connection to road safety measures introduced by legislators or regulators. None of the measures listed in Lord Robertson’s list of “we know what works” have been proven to work (Adams 1985, 1995).

In seeking an explanation for this paradox it will be helpful to place it in a wider risk management context. ► Figure 10.6 presents a risk typology that is germane to most discussions of a wide variety of risks and their management. Presented as a Venn diagram it suggests that it can be useful to distinguish three different, but not mutually exclusive, types of risk. Typing the single word “risk” into Google produces hundreds of millions of hits. One need sample only a small fraction in order to discover unnecessary and often acrimonious arguments caused by people using the same word to refer to different things and shouting past each



► Fig. 10.6  
Different kinds of risk

other. The typology offered in  Fig. 10.6 can help to dispose of some unnecessary arguments and civilize others.

Risks in the *perceived directly* circle are managed using judgment. We do not undertake a formal, probabilistic risk assessment before crossing the road; some combination of instinct, intuition, and experience usually sees us safely to the other side.

The second, *risk-perceived-through-science*, circle dominates the risk management literature. In this circle we find books, reports, and articles with verifiable numbers, cause-and-effect reasoning, probability, and inference. This is the domain of, amongst others, biologists with microscopes searching for microbial pathogens and astronomers with telescopes plotting the courses of incoming asteroids. This circle contains contributions from the whole range of science, technology, and the social sciences – from physics and chemistry to epidemiology and criminology. But the central science is statistics – the discipline that has probability at its core. This is the circle in which most of the published work on road safety can be found.

The circle labeled *virtual risk* contains contested hypotheses, ignorance, uncertainty, and unknown unknowns. If an issue cannot be settled by science and numbers we rely, as with directly perceptible risks, on *judgment*. Some find this enormously liberating; all interested parties feel free to argue from their beliefs, prejudices, or superstitions.

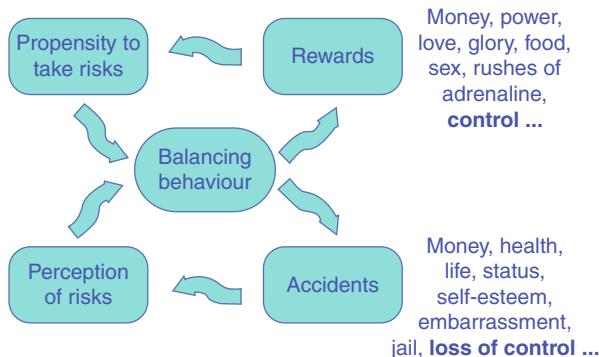
For students of risk this is by far the most challenging circle. The rules of mathematics and probability break down or there is insufficient data to invoke them. Although focused primarily on financial rather than physical risk, a provocative guide to behavior in this circle is Nassim Nicholas Taleb, author of *Fooled by Randomness* (2004) and *The Black Swan* (2007). On the road, as on the financial trading floor, all the participants react to the behavior of all the other participants. Risk behavior is reflexive and not well predicted by any known mathematical systems. It is in this circle that we find the longest-running and most acrimonious arguments. Virtual risks may or may not be real, but beliefs about them have real consequences.

Road safety is an intensively studied subject. It is awash with numbers, numbers adduced in support of the efficacy of existing road safety measures or in support of new ones proposed. And yet, despite all these numbers, we have the paradox described above. This suggests that most of the debate about road safety should be consigned to the third circle – virtual risk. After decades of road safety interventions we still appear to be unclear about what works.

Why should there remain any uncertainty about “what works”? Surely we know about the crash-protection benefits of seatbelts, air bags, and crumple zones. So why should the contribution of such benefits be so difficult to find in the aggregate statistical outcome?

## Different Risk Managers

A major part of the explanation lies in the fact that there are two very different sets of risk managers at work and their work tends to be found in different circles of  Fig. 10.6. One set might be called “institutional risk managers.” These are the legislators and regulators who make and enforce the rules governing transport safety, and the highway and vehicle engineers concerned with making roads and vehicles safer. Their quantitatively embellished work can be found mostly in the *perceived through science* circle. Their endeavors are routinely frustrated by the behavior of a second much larger set of risk managers consisting, worldwide, of billions of road users managing directly perceived risks guided by individual judgment.



**Fig. 10.7**  
The risk thermostat

Figure 10.7, the “risk thermostat,” presents a model of the risk management process that can help to demystify the paradox described above.

The model postulates that

- Everyone has a propensity to take risks – the setting of the thermostat.
- This propensity varies from one individual to another.
- This propensity is influenced by the potential rewards of risk taking.
- Perceptions of risk are influenced by experience of accident losses – one’s own and others’.
- Individual risk-taking decisions represent a balancing act in which perceptions of risk are weighed against propensity to take risks.
- Accident losses are, *by definition*, a consequence of taking risks (to take a risk is to do something that carries with it a probability of an adverse outcome); the more risks an individual takes, the greater, on average, will be both the rewards and the losses he or she incurs.

Credit for discovering this phenomenon is shared between a University of Chicago economist, Sam Peltzman (1975) after whom it is labeled by economists as the “Peltzman effect,” and a Canadian psychologist Gerald Wilde who dubbed it “risk compensation” and later “risk homeostasis.” Wilde’s most recent elaboration of the effect can be found in *Target Risk* (Wilde 1994, 2001).

The risk compensation model might also be called cost-benefit analysis without the £ or \$ signs. It describes a phenomenon known to the insurance industry as “moral hazard” – they have discovered that their customers are less careful about locking up if they have contents insurance. It is a conceptual model, not one into which you can plug numbers and from which you can extract decisions; the Rewards and Accidents boxes contain too many incommensurable variables. Our reasons for taking risks are many and diverse, and vary from culture to culture and person to person.

Most institutional risk managers work with a different model. “Reducing Risks, Protecting People” is the mantra of Britain’s Health and Safety Executive, the country’s foremost risk manager. It is also the title of the publication in which it explains its decision-making process (HSE 2001). In terms of Fig. 10.7, this process is confined to the bottom loop. It exemplifies the thought processes of most institutional risk managers, including those working on the

management of transport risks. Outside the offices of investment banks and hedge funds most institutional risk managers have only a bottom loop. Often their job specification precludes contemplation of the rewards of risk taking. Their job is to prevent accidents. The rewards loop is someone else's business – perhaps the marketing department.

But road users, whether pedestrians, cyclists, or motorists, have top loops. While trying to avoid accidents they are also in pursuit of the rewards of risk. These can range from getting from A to B on time, to the adrenaline rush of the boy racer or making contact with the person calling or texting one's mobile phone.

The model proposes that safety interventions that do not reduce the setting of the thermostat (propensity to take risks) will be offset by behavior that seeks to restore the balance of risk.

Antilock braking systems (ABS) provide a good example. When introduced, their superiority persuaded many insurance companies to offer discounts for cars with antilock brakes. Most of these discounts have now been withdrawn. The ABS cars were not having fewer accidents, they were having different accidents. Or perhaps they were having fewer accidents, but no fewer fatal accidents; the evidence from various studies is less than conclusive – leaving antilock brakes still in the disputed *virtual risk* category of Fig. 10.6.

The opening sentences of the Executive Summary of a recent US Department of Transport study on the long-term effect of ABS in passenger cars and LTVs states:

- ▶ Antilock brake systems (ABS) have close to a zero net effect on fatal crash involvements. Runoff-road crashes significantly increase, offset by significant reductions in collisions with pedestrians and collisions with other vehicles on wet roads. But ABS is quite effective in nonfatal crashes, reducing the overall crash-involvement rate by 6% in passenger cars and by 8% in LTVs (light trucks – including pickup trucks and SUVs – and vans) (NHTSA 2009).

The report notes that early studies of the initial effectiveness of ABS produced results that were “counterintuitive”:

- ▶ The overall effect of ABS on fatal crash involvements was close to zero.

Vehicles with four-wheel ABS had significantly higher rates of fatal run-off-road crashes than vehicles without ABS. In fact, the overall effect netted out to zero only because this increase was offset by a reduction in collisions with other vehicles on wet roads. These fairly strong statistical results did not square with intuition. The behavior of ABS on the test track did not provide any obvious reason that run-off-road crashes should increase; if anything, they suggested there ought to be a benefit.

In listing hypotheses to explain these perverse findings it is clear that the NHTSA's intuition was not informed by the risk compensation hypothesis. Still puzzled by their statistical findings, and seeking reassurance that antilock brakes are an effective safety measure, the report announces a 2008–2012 evaluation plan (Allen et al. 2008) that will seek to answer the following questions:

- What is the overall effect of ABS on nonfatal crashes?
- Even if the net effect of ABS on fatal crashes is close to zero, does ABS prevent enough nonfatal injuries and property damage to endorse ABS technology for its safety benefits? (p. 16)

It is sometimes argued that a risk compensation effect should only be found in cases where there is a clearly perceptible change in a vehicle's performance (IIHS 2007). It might help, it is accepted by some, to explain the statistical outcome associated with antilock brakes, but not

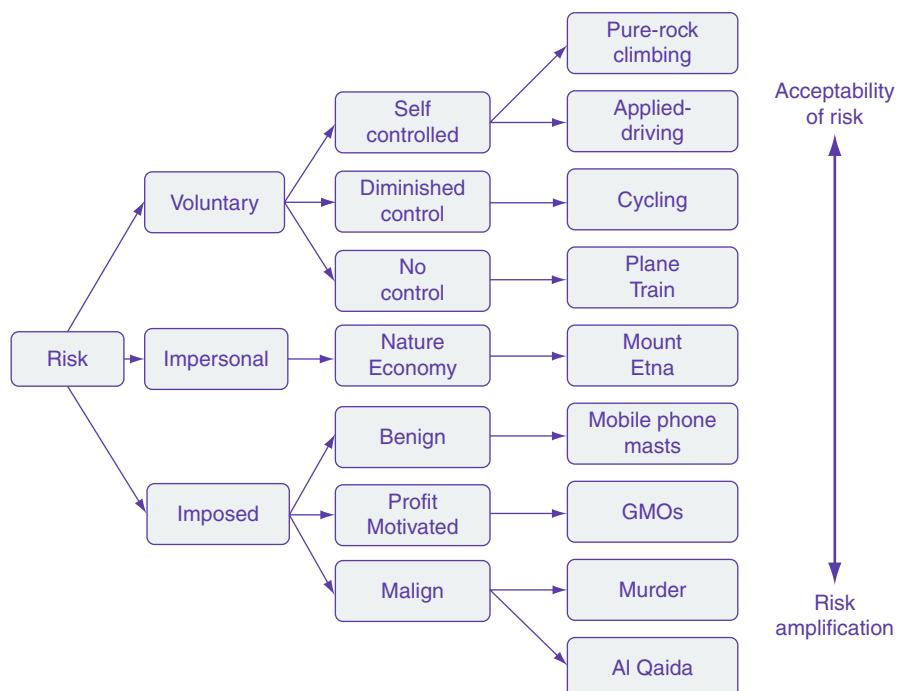
with seatbelts; that is, its effect should be confined to risks falling in the *directly perceptible* circle of the Venn diagram in [Fig. 10.6](#). But most people will admit to feeling safer when belted or, if habitual wearers of seatbelts, to feeling exposed and vulnerable without it. This feeling is surely amplified by highly publicized (and grossly exaggerated) claims for their effectiveness.

## What Kills You Matters

In listing some of the contents of the *Rewards* and *Accidents* boxes in [Fig. 10.7](#) control and loss of control were highlighted. [Figure 10.8](#) sets out the significance of this factor.

Acceptance of a given actuarial level of risk varies widely with the perceived level of control an individual can exercise over it and, in the case of imposed risks, with the perceived motives of the imposer.

- With “pure” voluntary risks, the risk itself, with its associated challenge and rush of adrenaline, is the reward. Most climbers on Mount Everest and K2 know that it is dangerous and willingly take the risk. Similarly thrill-seeking young men driving recklessly are aware that what they are doing is dangerous; that is the point.
- With a voluntary, self-controlled, applied risk, such as driving, the reward is getting expeditiously from A to B. But the sense of control that drivers have over their fates appears to encourage a high level of tolerance of the risks involved.



**Fig. 10.8**  
What kills you matters

- Cycling from A to B (I write as a London cyclist) is done with a diminished sense of control over one's fate. This sense is supported by statistics that show that per kilometer traveled a cyclist is much more likely to die than someone in a car. This is a good example of the importance of distinguishing between relative and absolute risk. Although much greater, the absolute risk of cycling is still small – 1 fatality in 25 million kilometers cycled; not even Lance Armstrong can begin to cover that distance in a lifetime of cycling. Numerous studies have demonstrated that the extra relative risk is more than offset by the health benefits of regular cycling; regular cyclists live longer.
- While people may voluntarily board planes, buses, and trains, the popular reaction to crashes in which passengers are passive victims suggests that the public demands a higher standard of safety in circumstances in which people voluntarily hand over control of their safety to pilots, or bus or train drivers.
- Risks imposed by nature – such as those endured by people living on the San Andreas Fault or the slopes of Mount Etna – or by impersonal economic forces – such as the vicissitudes of the global economy – are placed in the middle of the scale. Reactions vary widely. Such risks are usually seen as motiveless and are responded to fatalistically – unless or until the risk can be connected to base human motives. The damage caused by Hurricane Katrina to New Orleans is now attributed more to willful bureaucratic neglect than to nature. And the search for the causes of the economic devastation attributed to the “credit crunch” is now focusing on the enormous bonuses paid to the bankers who profited from the subprime debacle.
- Risks imposed by one's fellow humans are less tolerated. Consider mobile phones. The risk associated with the handsets is either nonexistent or very small. The risk associated with the base stations, measured by radiation dose, unless one is up the mast with an ear to the transmitter, is orders of magnitude less. Yet all around the world billions of people are queuing up to take the voluntary risk, and almost all the opposition is focused on the base stations, which are seen by objectors as impositions. Because the radiation dose received from the handset increases with distance from the base station, to the extent that campaigns against the base stations are successful, they will increase the distance from the base station to the average handset, and thus the radiation dose. The base station risk, if it exists, might be labeled a benignly imposed risk; no one supposes that the phone company wishes to murder all those in the neighborhood. The extent to which traffic is seen as an imposed risk varies widely. Parents of young children and cyclists are much more likely to feel it as an imposition than drivers of SUVs and big cars.
- Even less tolerated are risks whose imposers are perceived to be motivated by profit or greed. In Europe, big biotech companies such as Monsanto are routinely denounced by environmentalist opponents for being more concerned with profit than the welfare of the environment or the consumers of its products. Manufacturers of high-performance cars are assigned by some campaigners to the same category, their arguments sometimes adding damage to the environment to the danger posed to vulnerable road users.
- Less tolerated still are malignly imposed risks – crimes ranging from mugging to rape and murder. In most countries the number of deaths on the road far exceeds the numbers of murders, but far more people are sent to jail for murder than for causing death by dangerous driving. In the United States in 2002, 16,000 people were murdered – a statistic that evoked far more popular concern than the 42,000 killed on the road – but far less concern than that inspired by the zero killed by terrorists.

- This brings us to Al-Qaida and its associates. How do we account for the massive scale, worldwide, of the outpourings of grief and anger attaching to its victims, whose numbers are dwarfed by victims of other causes of violent death? In London, 52 people were killed by terrorist bombs on July 7, 2005, about 6 days worth of death on the road. But thousands of people do not gather in Trafalgar Square every Sunday to mark, with a 3-min silence, their grief for the previous week's road accident victims.

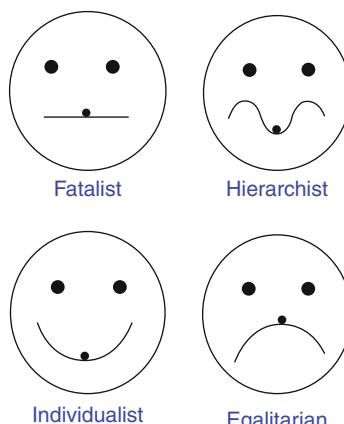
The dangers that can be tracked to the malign intent of terrorists are amplified by governments who see it as a threat to their ability to govern – to their ability to control events. To justify forms of surveillance and restrictions on liberty previously associated with tyrannies, “democratic” governments now characterize any risk to life posed by terrorists as a threat to *Our Way of Life*.

## Our Way of Life

How “we” manage risk to safeguard our way of life depends on who “we” are. ➤ [Figure 10.9](#) presents in cartoon form a typology of cultural biases commonly met in debates about risk (for the pre-cartoon version see Adams 1995).

These are caricatures, but nevertheless recognizable types that one encounters in debates about threats to safety and the environment. They have their origin in the work of anthropologist Mary Douglas on *cultural theory* (see Douglas and Wildavsky 1983; Douglas 1986). With a little imagination you can begin to see them as proponents and defenders of different ways of life. In a report for Britain’s Health and Safety Executive (Adams and Thompson 2002) they are described as follows:

- Individualists* are enterprising “self-made” people, relatively free from control by others, and who strive to exert control over their environment and the people in it. Their success is



■ **Fig. 10.9**  
A typology of cultural biases

often measured by their wealth and the number of followers they command. They are enthusiasts for equality of opportunity and, should they feel the need for moral justification of their activities, they appeal to Adam Smith's Invisible Hand that ensures that selfish behavior in a free market operates to the benefit of all. The self-made Victorian mill owner or present-day venture capitalist would make good representatives of this category. They oppose regulation and favor free markets. Nature, according to this perspective, is to be *commanded* for human benefit.

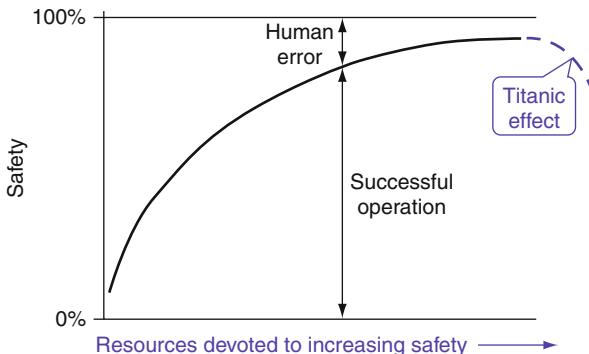
- *Egalitarians* have strong group loyalties but little respect for externally imposed rules, other than those imposed by nature. Human nature is – or should be – cooperative, caring, and sharing. Trust and fairness are guiding precepts and equality of outcome is an important objective. Group decisions are arrived at by direct participation of all members, and leaders rule by the force of their arguments. The solution to the world's environmental problems is to be found in voluntary simplicity. Members of religious sects, communards, and environmental pressure groups all belong to this category. Nature is to be *respected and obeyed*.
- *Hierarchists* inhabit a world with strong group boundaries and binding prescriptions. Social relationships in this world are hierarchical with everyone knowing his or her place. Members of caste-bound Hindu society, soldiers of all ranks, and civil servants are exemplars of this category. The hierarchy certifies and employs the scientists whose intellectual authority is used to justify its actions. Nature is to be *managed*.
- *Fatalists* have minimal control over their own lives. They belong to no groups responsible for the decisions that rule their lives. They are nonunionized employees, outcasts, refugees, untouchables. They are resigned to their fate and see no point in attempting to change it. Nature is to be *endured* and, when it is your lucky day, *enjoyed*. Their risk management strategy is to buy lottery tickets and duck if they see something about to hit them.

Transport risk managers, in the terms of this typology are statuary Hierarchists who make the rules and enforce the rules. For the foreseeable future they can expect to be attacked from the Egalitarian quadrant for not doing enough to protect us, and from the Individualist quadrant for over regulating and suffocating freedom and enterprise.

During the public debate that preceded the passage of Britain's seatbelt law the principal participants could be readily assigned to quadrants of this typology. The proponents of the law were Hierarchists, otherwise labeled "the Nanny State" by Individualists in the lower left-hand quadrant who were, in turn, labeled "loony Libertarians" by the law's supporters. The Egalitarian quadrant was divided. It contained traditional safety campaigners such as the Royal Society for the Prevention of Accidents and the Parliamentary Advisory Council on Transport Safety who supported the law. But it also contained campaigners for pedestrian and cycling safety who had bought into what was then the radical new idea of risk compensation and saw the seatbelt law as a threat to their constituents.

## What Sort of Risk?

- *Figure 10.10* below, borrowed (and amended) from the risk management manual of a major airline, presents yet another way of looking at the different types of risk set out in *Fig. 10.6*.



**Fig. 10.10**  
The human reliability curve

On the steep part of the curve risks, whether up Mount Everest or down a Victorian (or Chinese) coal mine, are usually obvious (directly perceptible), but the responses are diverse and often contentious. Certainly traditional Everest mountaineers are resentful of bureaucratic interference in their risk taking. But their traditions are being compromised by commercial tour companies who, at great expense, offer to guide people to the top and back *safely*. A fatality in 1999 led to a claim of negligence and an out-of-court settlement for £70,000 (Mountain Clients 2007). Britain's oxymoronic Adventure Activities Licensing Authority, instituted to ensure safe adventure, is also seen, by Individualists, as a threat to traditional risk-taking freedoms.

Large risks associated with employment are commonly viewed as imposed risks, imposed by economic necessity especially when the employees are poor (as in the case of Victorian, or Chinese, coal miners). Here interventions in the form of regulation and inspection are more readily accepted – but not always with the expected result. The Davy Lamp (named after its inventor Sir Humphry Davy), which most histories of science and safety credit with saving thousands of lives, is usually described as one of the most significant safety improvements in the history of mining. But it appears to have been a classic example of a potential safety benefit consumed as a performance benefit. Because the lamp operated at a temperature below the ignition point of methane, it permitted the extension of mining into methane-rich atmospheres; the introduction of the “safety lamp” was followed by an increase in explosions and fatalities (Albury and Schwarz 1982).

But when all the obvious measures are in place accidents will still, occasionally, happen. Hundred percent safety is a utopian goal. Indeed it is possible to have too many safety measures. So long as there is a residual dependence on the vigilance of fallible humans, their level of vigilance will depend on the strength of their belief that something can go wrong. The impressive safety record of civil aviation, and all the safety redundancy built into modern aircraft and their operating systems have created a problem of keeping pilots awake on long flights across time zones. Why should they stay alert for the whole of their working lives in anticipation of something they believe will never happen? When you are on the flat part of the curve you do not have a clue whether further safety precautions will have any beneficial effect. The area above the flat part of the *human reliability curve* might be

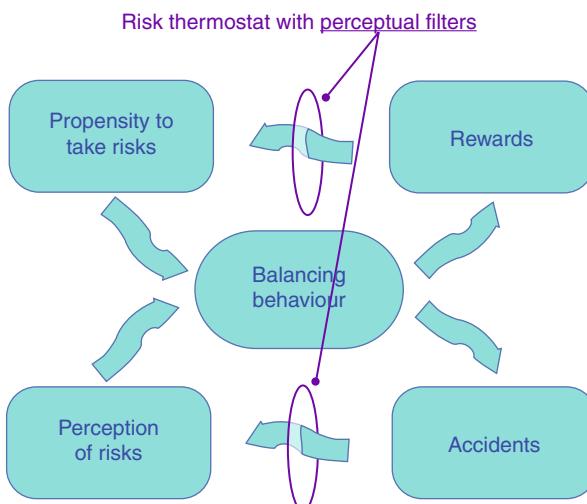
thought of as a zone of *virtual risk*. There are circumstances within this zone where further safety measures can have a perverse effect – where the belief in such measures can induce complacency – the *Titanic Effect*.

## Filters

The variety to be found in the risks and rewards boxes of the *risk thermostat* and the variable responses to them illustrated in [Fig. 10.8](#) (what kills you matters) by the different actors presented in [Fig. 10.9](#) suggest that the *risk thermostat* should be fitted with perceptual filters. The same objective facts can have an enormously varied influence on risk-taking behavior.

[Figure 10.11](#) can serve as a description of the behavior of the driver of a single car going around a bend in the road. His speed will be influenced by his perception of the rewards of risk; these might range from getting to the church on time to impressing his friends with his skill or courage. His speed will also be influenced by his perception of the danger; his fears might range from death, through the cost of repairs and loss of his license, to mere embarrassment. His speed will also depend on his judgment about the road conditions – Is there ice or oil on the road? How sharp is the bend and how high the camber? – and the capability of his car – How good are the brakes, suspension, steering, and tires?

Overestimating the capability of the car or the speed at which the bend can be safely negotiated can lead to an accident. Underestimating those things will reduce the rewards gained. The consequences, in either direction, can range from the trivial to the catastrophic. The balancing act described by this illustration is analogous to the behavior of a thermostatically controlled system. The setting of the thermostat varies from one individual to another, from one group to another, from one culture to another, and for all of these, over time. Some like it



**Fig. 10.11**  
Perceptual Filters

hot – a Hell's Angel or a Grand Prix racing driver, for example – others like it cool – a Casper Milquetoast or a little old lady named Prudence. But no one wants absolute zero.

## Risk: An Interactive Phenomenon

Figure 10.12 introduces a second road user to make the point that risk is usually an interactive phenomenon. One person's balancing behavior has consequences for others. On the road one motorist can impinge on another's "rewards" by getting in their way and slowing them down, or help them by giving way. One is also concerned to avoid hitting other motorists or being hit by them. Driving in traffic involves monitoring the behavior of other motorists, speculating about their intentions, and estimating the consequences of a misjudgment. Drivers who see a car approaching at high speed and wandering from one side of the road to the other are likely to take evasive action, unless perhaps they place a very high value on their dignity and rights as a road user and fear a loss of esteem if they are seen giving way. During this interaction enormous amounts of information are processed. Moment by moment each motorist acts upon information received, thereby creating a new situation to which the other responds.

On the road and in life generally, risky interaction frequently takes place on terms of gross inequality. The damage that a heavy truck can inflict on a cyclist or pedestrian is great; the physical damage that a cyclist or pedestrian might inflict on the truck is small. The truck driver in this illustration can represent the controllers of large risks of all sorts. Those who make the decisions that determine the safety of consumer goods, working conditions, or large construction projects are, like the truck driver, usually personally well insulated from the consequences of their decisions. The consumers, workers, or users of their constructions, like the cyclist, are in a position to suffer great harm, but not inflict it.

The world, at the time of writing, contains about 6.5 billion risk thermostats, and they interact. Figure 10.13, the *Dance of the Risk Thermostats*, provides a tiny window on a few of these interactions. Some of the thermostats are large – presidents with fingers on buttons – most are tiny – shepherds in Afghanistan or children chasing balls across streets. In a rapidly globalizing world the lines of interaction are growing longer and more numerous.

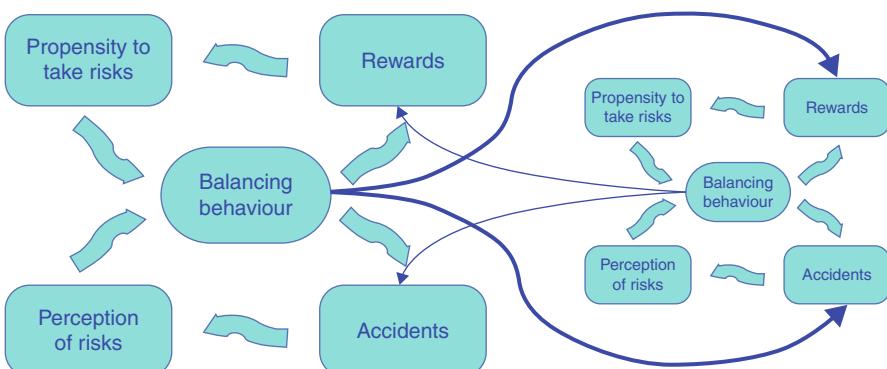


Fig. 10.12  
The truck driver and the cyclist

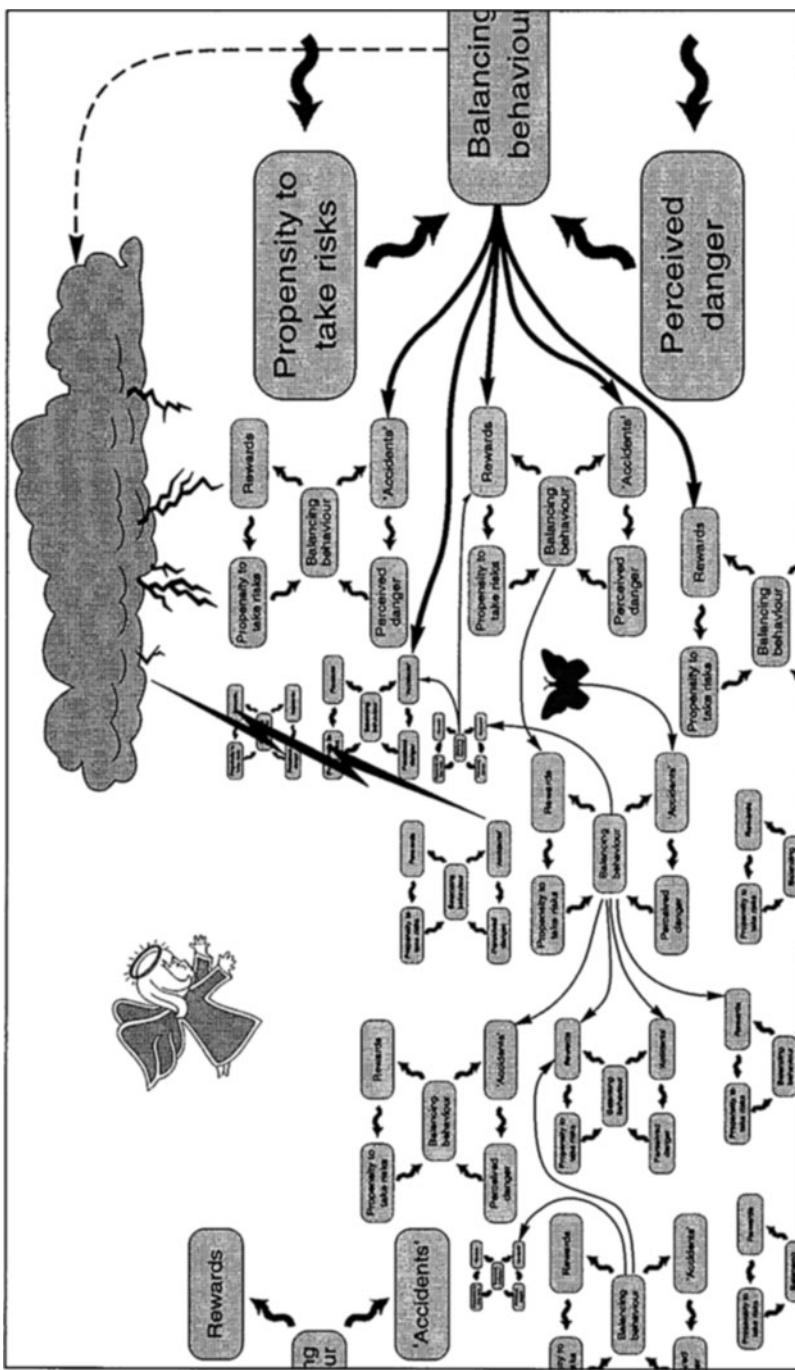


Fig. 10.13  
Dance of the risk thermostats

Overhanging everything are the sometimes destructive forces of nature – droughts, floods, earthquakes, hurricanes, and plagues. The broken line symbolizes the uncertain impact of human behavior on nature. Lurking below are those seeking to control or influence them, from would-be climate engineers to shamans conducting rain dances. Fluttering about the dance floor are the Beijing butterflies beloved of chaos theorists: They ensure that the best laid plans of mice and men “gang aft agley.”

The winged creature at the top left was added in response to a survey that revealed that 69% of Americans believe in angels and 46% believe they have their own guardian angel. The “angel factor” must influence many risk-taking decisions – from those of suicide bombers to those of risk-taking motorists; Deus é Brasileiro (God is Brazilian) is an expression invoked by Brazilian motorists whose wild driving under purported divine protection has terrified me (Adams 2009, Preface).

❸ [Figure 10.13](#) shows but an infinitesimal fraction of the possible interactions between all the world’s risk thermostats; there is not the remotest possibility of ever devising a model or building a computer that could predict accurately all the consequences of intervention in this system.

In the mix are jihadists and CIA operatives, financial regulators and subprime mortgage brokers, occupational health and safety regulators, employers and employees, doctors, no-win-no-fee lawyers, police judges, and juries. And in the realm of transport risks one finds engineers, regulators, and the regulated.

## Conclusion

---

Most transport risks are likely to remain in the contested *virtual risk* circle of ❸ [Fig. 10.6](#). Many will doubtless continue to insist that they know what works. However the United Nations Decade of Action for Road Safety and the Make Roads Safe campaign referred to at the beginning of this essay would appear doomed to disappointment in the developing countries on which their efforts are focused. Wherever one looks one finds the tendency illustrated by ❸ [Figs. 10.1–10.3](#) repeated. As the number of cars in a country increases, the death rate per car decreases. In countries in the early stages of motorization each vehicle is incredibly lethal. Poor countries with a small number of modern cars, with 100 years of safety technology built into them, are achieving kill-rates per vehicle as high or higher than those at the time of Model-Ts. This phenomenon has become known as Smeed’s Law, after Reuben Smeed who established the relationship over 50 years ago (Adams 1985, 1987).

The confidence of some institutional transport safety managers that they “know what works” is undermined all round the world by the behavior of billions of individual risk managers who *react* to the impositions of the official risk managers, but also to the behavior of everyone else on the road. It is known that seatbelts provide significant protection in crashes, that helmets reduce injury caused by a knock on the head, that antilock brakes are superior brakes, that alcohol increases the likelihood of accidents, and speed their severity. But whenever safety measures attempting to put this knowledge to effective use are imposed from on high by institutional risk managers, the result is at best disappointing.

So what did cause the declining death rates described by the Smeed Law? Here we must speculate; the myriad interactions involved in the dance of the risk thermostats defy capture by any known computer. If one accepts ❸ [Fig. 10.7](#) as a plausible description of the process of risk

management, one looks to changes in the setting of the thermostat for an explanation. As we get richer we become more risk averse.

Car ownership correlates strongly and positively with income. As nations become richer they can afford, *and demand*, higher levels of safety and security. The setting of the collective thermostat is turned down. Reference was made above to the risks experienced in Victorian (or Chinese) coalmines. In poor countries life is cheaper and safety standards of all sorts are lower; life expectancy at birth is much lower and road accident rates much higher.

In the most affluent countries of the world there is a trend toward increasing institutional risk aversion and growth in the numbers of institutional risk managers. Their job is to reduce accidents, and then get them lower still. For them, one accident is one too many. As noted above their risk thermostats have no top loop. But despite the increase in the activity of institutional risk managers it is often difficult to discern the effect of their work. As in the case of “The Scandinavian Myth” discussed above, their growing activity appears to be symptomatic of increasing societal risk aversion rather than the cause of a decrease in accidents.

Growing concern for the safety of children on the road might serve as another indicator of an increase in societal risk aversion in affluent countries, and an explanation for a significant part of the plummeting death rate illustrated in [Figs. 10.1](#) and [10.2](#). Today in Britain, per 100,000 children, the road accident death rate is less than a quarter of what it was in 1922 when there was hardly any motorized traffic and the country had a nationwide 20 mph speed limit. This is not because the streets have become safer for children to play in; there is now much more metal in motion. It is because few children are allowed out on their own anymore. In 1971, 80% of 7- and 8-year-old children got to school unaccompanied by an adult. By 1990, this had dropped to 9% (Hillman et al. [1990](#)), and by 2010 it had become a child protection issue (in England in 2010, two controversies appeared in the press in which parents were threatened with child protection orders for allowing their children what used to be the widely accepted freedom to get to school unaccompanied: <http://www.telegraph.co.uk/family/7872970/Should-the-Schonrock-children-be-allowed-to-cycle-to-school-alone.html> and <http://www.bbc.co.uk/news/uk-england-lincolnshire-11288967>). The decrease in child road accidents appears to be overwhelmingly attributable to a decrease in exposure, and the decrease in exposure attributable not to institutional edict but to a growing fear on the part of parents of the threat posed to their children by traffic.

At present the two countries with the best road safety records in the world are pursuing diametrically opposed philosophies of road safety. The Swedish “Vision Zero” policy assigns ultimate responsibility for road safety to the institutional risk manager in the form of the state. The responsibility of users of the system is to obey the rules. It asserts that the rules for the system are that:

1. The designers of the system are always ultimately responsible for the design, operation, and use of the road transport system and thereby responsible for the level of safety within the entire system.
2. Road users are responsible for following the rules for using the road transport system set by the system designers (e.g., wearing seat belts and obeying speed limits).
3. If road users fail to obey these rules due to lack of knowledge, acceptance, or ability, or if injuries occur, the system designers are required to take necessary further steps to counteract people being killed or seriously injured (Hill [2008](#)).

In the Netherlands, a country with an even (slightly) better road safety record, there is a growing enthusiasm for “shared space.” This is an intriguing idea pioneered by the late Hans Monderman, a highway engineer in Friesland. He removed almost all the traffic lights, pedestrian barriers, stop signs, and other road markings that had been assumed to be essential for the safe movement of traffic.

For traditional highway engineers his idea was anathema. Since the advent of the car they have planned on the assumption that car drivers are selfish, stupid, but obedient automatons who had to be protected from their own stupidity, and that pedestrians and cyclists were vulnerable, stupid, obedient automatons who had to be protected from cars – and their own stupidity. Hence the ideal street was one in which the selfish-stupid were completely segregated from the vulnerable-stupid, as on the American freeway or European motorway where pedestrians and cyclists and pedestrians are forbidden. Where segregation was not possible, in residential suburbs and older urban areas, their compromise solution was the ugly jumble of electronic signals, stop signs, barriers, and road markings that now characterize most urban environments.

Monderman observed those using the streets for which he was responsible and concluded that they were not stupid, but neither did they obey all the rules and barriers that assumed that they were nor, on the whole, did they behave selfishly. Pedestrians, he noticed, were nature’s Pythagoreans – always preferring the hypotenuse to the other two sides of the triangle. Given half a chance they did not march to the designated crossing point and cross at right angles to the traffic; if they spotted a gap in the traffic they opted for the diagonal route of least effort.

And motorists did not selfishly insist on their right of way at the cost of mowing down lots of pedestrians. Monderman decided that those for whom he was planning were vigilant, responsive, and responsible. He deliberately injected uncertainty into the street environment about who had the right of way. The results were transformative. Traditional highway engineers have never been concerned with aesthetics. Their job was to move traffic safely and efficiently. They dealt not with people but PCUs (passenger car units). The removal of the signals, signs, and barriers that were the tools of their trade not only greatly improved the appearance of the streetscape but, by elevating the status of the pedestrian and cyclist relative to that of the motorist, made them more convivial as well. Monderman was a practitioner, and so far as I am aware published nothing in the peer-reviewed literature, but his practice has been enormously influential (see Wikipedia – [http://en.wikipedia.org/wiki/Hans\\_Monderman](http://en.wikipedia.org/wiki/Hans_Monderman) – and numerous other online references).

Claes Tingvall, who is credited with being the architect of Sweden’s Vision Zero, said in an interview “Vision Zero . . . is a shift in philosophy. Normal traffic policy is a balancing act between mobility benefits and safety problems. The Vision Zero policy refuses to use human life and health as part of that balancing act; they are non negotiable. . . . Part of the Vision Zero strategy is to improve the demand for safety” (Tingvall 2009).

## A Concluding Speculation

---

Tingvall’s characterization of “normal traffic policy” as “a balancing act between benefits and safety” is a fair approximation of the risk management behavior described by the risk thermostat in Fig. 10.7. But who decides that the thermostat should be set to zero? If it were truly set to zero for all road users no one would move.

In both Sweden and the Netherlands, one senses a high and still growing demand for safety. This is perhaps the ultimate explanation of the good accident records of both. This increasing demand might be characterized as a progressive reduction of the setting of both the Dutch and the Swedish societal risk thermostats.

Every pedestrian, cyclist, and motorist is also a risk manager, performing “a balancing act between benefits and safety.” Anyone with direct experience of how this act is performed in countries at the early stages of motorization (as well as those studying their accident statistics) will know that the performance in such countries is very different from that in highly motorized countries.

The UN’s Decade of Action for Road Safety seeks to promote road safety everywhere but is focused primarily on the least motorized countries with the highest accident rates. The claim quoted at the beginning of this essay that “we know what works,” in the light of the evidence reviewed here, appears hubristic. Unless and until ways are devised to lower the settings of the collective risk thermostats of these countries, the slaughter on their roads looks destined to increase in the early rapid-growth stage of their motorization. The policy maker’s choice of setting of the thermostat is of marginal relevance; it is the average setting of the thermostats of *all* the participants in complex interactive systems that determines the accident outcomes.

The challenge for those trying to make roads safer is to change attitudes – to promote greater risk aversion on the road. There are encouraging precedents. The stigmatizing of smoking and drunken driving has greatly reduced the practice of both; but the change took many years. Perhaps the widespread distribution of Make-Roads-Safe T-shirts, banners, publications, and wristbands by the Make Roads Safe campaign is not a bad way to start. Recruitment of the endorsement of the campaign by superstars of Formula 1, the most spectacular possible exemplars of high-risk driving, is a less obvious method of promoting risk aversion on the road.

## References

---

- Adams J (1985) Risk and freedom: the record of road safety regulation transport publishing projects. <http://john-adams.co.uk/wp-content/uploads/2007/10/risk%20and%20freedom.pdf>
- Adams J (1987) Smeed’s law: some further thoughts. *Traffic Eng Contr* 28(2):70–73
- Adams J (1995) Risk. UCL Press, London
- Adams J (2009) Risco, Editora Senac, São Paulo. The preface, Deus é Brasileiro, is available in English. <http://john-adams.co.uk/wp-content/uploads/2008/12/deus-e-brasileiro1.pdf>
- Adams J (2010) Seatbelts. <http://john-adams.co.uk/category/seat-belts/>
- Adams J, Thompson M (2002) Taking account of societal concerns about risk: framing the problem, research report 035. Health and Safety Executive. Available online at <http://www.hse.gov.uk/research/rrpdf/rr035.pdf>
- Albury D, Schwarz J (1982) Partial progress. Pluto Press, London
- Allen K, Dang JN, Doyle CT, Kahane CJ, Roth JR, Walz MC (2008) Evaluation program plan, 2008–2012. NHTSA technical report, DOT HS 810 983. National Highway Traffic Safety Administration, Washington, DC, pp 1, 9
- Avineri E et al (2009) Individual behaviour change: evidence in transport and public health, department for transport contract PPRO 04/06/33, project team: Erel Avineri, Kiron Chatterjee, Andrew Darnton, Phil Goodwin, Glenn Lyons, Charles Musselwhite, Paul Pilkington, Geof Rayner, Alan Tapp, E. Owen D. Waygood, and Peter Wiltshire
- Department for Transport (2008) 25th anniversary of seatbelts – 60,000 lives saved (press release). [http://www.direct.gov.uk/en/N11/Newsroom/DG\\_072333](http://www.direct.gov.uk/en/N11/Newsroom/DG_072333)
- Douglas M, Wildavsky A (1983) Risk and culture: an essay on the selection of technological and environmental dangers. University of California Press, Berkley
- Douglas M (1986) Risk acceptability according to the social sciences. Routledge and Kegan Paul, London

- Haight FA (1985) The developmental stages of motorization: implications for safety. Pennsylvania Transportation Institute, State University, Penn
- Health and Safety Executive (2001) Reducing risks, protecting people: HSE's decision-making process, Crown copyright 2001. <http://www.hse.gov.uk/risk/theory/r2p2.pdf>
- Hill J (2008) Getting ahead: returning Britain to European leadership in road casualty reduction (PDF). Road safety foundation. <http://www.saferoaddesign.com/media/1752/bookletweb.pdf>
- Hillman H et al (1990) One false move a study of children's independent mobility. Policy Studies Institute, London
- HLDI (Highway Loss Data Institute) (2009) Hand-held cellphone laws and collision claim frequencies. Loss Bull 26(17). Arlington, VA. [http://www.iihs.org/research/topics/pdf/HLDI\\_Cellphone\\_Bulletin\\_Dec09.pdf](http://www.iihs.org/research/topics/pdf/HLDI_Cellphone_Bulletin_Dec09.pdf).
- HLDI (2010) Texting laws and collision claim frequencies. HLDI Bull 27(11). ([http://www.iihs.org/research/topics/pdf/HLDI\\_Bulletin\\_27\\_11.pdf](http://www.iihs.org/research/topics/pdf/HLDI_Bulletin_27_11.pdf))
- IIHS (2007) Risk compensation keeps popping up where it's totally irrelevant, insurance institute for highway safety. IIHS Advisories, No. 33. [http://www.iihs.org/research/advisories/iihs\\_advisory\\_33.html](http://www.iihs.org/research/advisories/iihs_advisory_33.html)
- Law Commission (Ireland) (2008) Statute law repeals: consultation paper city of Dublin repeal proposals. <http://www.lawcom.gov.uk/docs/dublin.pdf>
- Make Roads Safe (2010) Make roads safe is the campaign for global road safety. <http://www.makeroadssafe.org/about/Pages/homepage.aspx>
- Mountain Clients (2007) Family launches lawsuit over Everest death. <http://www.mountain-clients.org.uk/2008/06/family-launches-lawsuit-over-everest-death-23-jan-2007/>
- NHTSA (2009) The long-term effect of ABS in passenger cars and LTVs DOT HS 811 182. [www-nrd.nhtsa.dot.gov/Pubs/811182.pdf](http://www-nrd.nhtsa.dot.gov/Pubs/811182.pdf)
- Parliamentary Advisory Council for Transport Safety (2010) PACTS newsletter: DfT celebrates 25th anniversary of the introduction of seatbelts. <http://www.pacts.org.uk/newsletters.php?id=2>
- Peltzman S (1975) The effect of automobile safety regulation. *J Polit Econ* 83(4):677–725
- Plowden W (1971) The motor car and politics 1896–1970. Bodley Head, London, p 455
- Ross HL (1976) The Scandinavian myth: the effectiveness of drinking and driving legislation in Sweden and Norway. In: Evaluation studies – review annual, vol 1. Sage, London
- Royal Society for the Prevention of Accidents (2010) Our success stories. <http://www.rospa.com/about/successstories/>
- Taleb NN (2004) Fooled by randomness: the hidden role of chance in life and in the markets. Random House, New York
- Taleb NN (2007) The Black Swan: the impact of the highly improbable. Penguin, London
- Tingvall C (2009) Road safety interview: Sweden's vision zero, Allianz. [http://knowledge.allianz.com/mobility/road\\_safety/?451/road-safety-vision-zero](http://knowledge.allianz.com/mobility/road_safety/?451/road-safety-vision-zero)
- UN Department of Public Information (2010) General assembly adopts text proclaiming decade of action for road safety (2011–2020), aimed at reducing traffic-related deaths, injuries. <http://www.un.org/News/Press/docs/2010/ga10920.doc.htm>
- Vinglis ER (1983) Guest editor's introduction. *Accid Anal Prev* 15:405–406
- USDoT (2010) Make no mistake; DOT and its safety partners will continue fighting against distracted driving. <http://fastlane.dot.gov/2010/09/make-no-mistake-dot-and-its-safety-partners-will-continue-fighting-against-distracted-driving.html>
- Wilde GS (1994) Target risk. PDE Publications, Toronto. <http://psyc.queensu.ca/target/#contents>
- Wilde GS (2001) Target risk 2, 2nd edn. PDE Publications, Toronto

# 11 Risk and Spatial Planning

Claudia Basta

Delft University of Technology, 3TU. Centre for Ethics and Technology,  
Delft, The Netherlands  
Wageningen University and Research Centre, Wageningen,  
The Netherlands

<b>Introduction .....</b>	<b>267</b>
The “Siting Risks” Issue .....	270
<b>History .....</b>	<b>272</b>
<b>Current Research .....</b>	<b>275</b>
Finding a Place to Risks: The Example of the European Regulatory Framework on Dangerous Substances .....	275
Deciding on Chances or Deciding on Consequences? European Answers and Their Ethical Basis .....	278
Framing Siting Controversies: The Case of the CO <sub>2</sub> Storage in Barendrecht, The Netherlands .....	283
Safety as a Spatial Value: Siting Risks as a Matter of Distributive Justice .....	286
Concluding Remarks: Toward a Moral Understanding of the Relation Between Risks and Space .....	290
<b>Further Research .....</b>	<b>291</b>

**Abstract:** Proponents of site-specific hazardous technologies and members of involved communities are often in conflicting positions regarding the most appropriate location for their siting. Because of the component of uncertainty that characterizes the assessment of the potential consequences of these technologies and the different perception of risks by the side of individuals, the “where of risks” is rarely uncontroversial.

This chapter discusses the relation between “risks” and “space” and argues, in particular, on its moral implications. Such implications regard the land use planning evaluations related to the risks (e.g., of the release of hazardous substances or radioactive emissions) arising from these technologies. These risks constitute the main locational criteria. This chapter reflects on the moral legitimacy of the development and outcomes of locational assessments by arguing on possible forms of synergy between spatial planning theories and ethical theories.

In the first part of this chapter, a concrete example of a European chemical safety regulation (namely, Directive 96/82/EC on Hazardous Substances, the so-called Seveso Directive) is discussed. Article 12 of the Directive, that is, the “Control of Urbanization” requirement, requires member states to assess and maintain opportune safety distances from Seveso establishments according to the risk of major accidents. This requirement is of particular interest in the context of this chapter as it offers the opportunity to investigate different methods used to perform the assessment of safety distances that are “spatially relevant” safety measures. Such methods are implemented in selected European countries. Their differences relate not only to considerations of technical nature but also to different cultural attitudes toward risk. Somehow there are also non-explicit, although identifiable different ethical assumptions at their basis. These assumptions and their implications will be therefore discussed.

The second part of this chapter focuses on the matter of framing the “siting risks” issue correctly. The matter of overcoming the predominant tendency of interpreting all siting controversies as “not-in-my-backyard” (in the following: NIMBY) situations will be discussed in detail. It is proposed that controversies that do evidently lose, along their development, identifiable spatial and temporal dimensions may signal a general social opposition to a given technological development rather than a merely local rejection of a technological installation. This distinction between site-specific and “a-site” controversies is particularly important for both the planning theory and practice. Spatial planning processes are the appropriate framework for addressing and solving NIMBY cases; however, they become merely instrumental in those cases in which it is not the “where” but the “if” of the technology to be object of discussion. As it will be discussed the moral implications to be considered in the two different circumstances are of remarkably different nature. Arguing on the distinctive elements of NIMBY versus non-NIMBY cases does therefore occupy this part.

Having presented some examples of European national approaches to the siting of hazardous installations and having provided the distinctive elements of the cases to be framed within a spatial planning discourse, the final part of this chapter concentrates on the possible integration of ethical and spatial planning theories. By referring to the Rawlsian theory of justice (1971) as applied to spatial planning theories by Moroni (1997), a conception of “spatial safety” as primary spatial good is proposed. By taking this perspective aim of the spatial planning practice in a fair society becomes distributing spatial safety equally up to the lowest societal level. As it will be argued this theoretical approach to the distribution of spatial safety in society has important, and promising, implications particularly for the planning practice. First, it implies an evaluative shift from the single siting case to the higher regional or national scale; second, next to spatial and risk tolerability criteria it provides the planner with a *moral*

criterion (i.e., the fair distribution of risks in society) for justifying the outcomes of locational assessments. Indications for the further research needed to strengthen this fruitful integration of risk, spatial planning, and ethical theories in the context of hazardous facility siting are provided in the conclusive part of this chapter.

## Introduction

---

Driving on a highway in Northern Europe, contemplating the countryside and being suddenly adjacent to a large industrial site is a rather different experience in comparison to entering the historical city of Venice through the Liberty Bridge and looking toward the lagoon across the fumes of the petrochemical harbor. However, these experiences have something in common: the evident background of technology and the invisible presence of risks.

If also these risks could be visualized they would appear as potential areas of impacts within which our life and health, together with those of those around us, are under threat. The threats consist of the possible release of hazardous substances, of noxious radiations, or of the sudden wave of overpressure caused by an explosion. If these scenarios could be visualized while we experience the living environment, what we would see is that, simply, we are daily *at risk*.

Among the hazardous technologies coloring the geographical maps designed by risk analysts to visualize potential areas of impacts there are nuclear power plants, LPG storages, chemical establishments, hydrogen-fuel stations, and generally all installations that, while providing sometimes vital benefits to society, have the potential to harm. Despite they do not involve any form of industrial processing and are underground, emerging technologies such as carbon dioxide disposals (in the following: CO<sub>2</sub> disposals) could be also accounted. The common feature among these technologies is being site-specific and posing risks to the man and the environment because of a specific hazardous factor. To prevent and minimize the consequences of the deriving risks, these technologies must therefore be object of a careful locational assessment.

It is rather banal stating that a chemical factory should not be placed within a residential district or a nuclear power plant in a highly instable hydrogeological area. Nevertheless, locational assessments are much more complex than what common sense would lead one to conclude. Finding the most appropriate site to risks consists of a difficult search of balance among land scarcity, an unachievable risk-free living environment, and the many material and immaterial benefits contemporary society opted to rely upon. That is why hazardous facilities siting processes are rarely uncontroversial. Risky installations stigmatize localities and affect the identity of places as well as individuals in a sometimes irreversible way (Boholm and Löfstedt 2004). Loss of property values, phenomena of industrial encroachments, spoiled place-, and self-perception are recurrently documented implications of living nearby hazardous facilities (Boholm and Löfstedt 2004; Lesbirel and Shaw 2005). As periodically reported by the media policy-makers, proponents of such installations and the public are often polarized when decisions about their siting have to be taken.

This problem is well known in literature and in the past two decades it has been object of increasing attention by the side of a variety of scholarly fields. “Siting risks” implies accounting a large number of factors, which are mirrored by the many disciplinary fields that provided their contributions to what has, in time, become an established literature. Considering the

multidimensionality of the concept of risk and its technology-specific and site-specific implications, the siting of risk-bearing installations can be considered a problem of “irreducible complexity” (Renn 2006). Despite the acknowledged difficulty of approaching such complexity through a generally applicable theoretical framework, recent research explored the cross-technology, cross-boundaries, and common ethical questions that characterize facility siting processes beyond their more specific implications. In particular, in her editorial on the new perspectives on siting controversies Boholm (2004) identifies three main areas of investigations that, she suggests, may help in overcoming exploited interpretational shortcuts and framing facility siting conflicts correctly. One of these shortcuts is interpreting all siting controversies as not-in-my-backyard (in the following: NIMBY) situations. “NIMBY” is the popular label attached to those siting controversies in which more or less fierce oppositions polarize risk posers and risk runners. In the majority of cases, the NIMBY perspective sees the former as the promoters of the general good and the latter as the selfish defenders of local interests. Under this light siting controversies are reduced to oppositions not only between polarized actors, but also between a technocratic and neutral versus an irrational and biased appraisal of the risks at issue. Somehow, the NIMBY dispute becomes a confrontation between rationality and irrationality.

The underlying assumption of this predominant interpretation of siting controversies is that a motivated collective interest finds an obstacle in a selfish local resistance. The limitedness of this interpretation is quite evident; by reducing all siting controversies to these minimal terms, the overall desirability of the technology under consideration, its effective compatibility with the chosen context, and the legitimacy of the decision-making process promoting its installation are not the real focus of the discussion. What it is so is rather preventing the supposed collective benefits from finding an obstacle in narrow-minded forms of rejections. Empowering planning procedures to enforce siting processes regardless of these rejections while limiting their duration in time becomes, consequently, the real focus of decisional processes and furthermore policy making (Owens 2004). Privileged instruments in this respect are risk communication strategies that delegitimize subjective and emotional appraisals of risks and the introduction of forms of compensation for at-risk communities.

Inevitably, the legitimacy of such “enforcing instruments” has been put under severe discussion by risk scholars. Both metaethical studies on their assumptions (Roeser 2006, 2010) and interdisciplinary studies on their ethical implications (Linnerooth-Bayer, in Lesbirel and Shaw 2005) denounced their possible fallacies. There is growing consensus that throughout their development, siting processes require a rigorous consideration of their moral implications and that the interests and perceptions of involved parties should be simultaneously accounted. On the one hand, “there are no universal *norms* on which to base a siting strategy” (Linnerooth-Bayer 2005, in Lesbirel and Shaw 2005, at 59), as the perceptions and interests of involved parties do strongly depend on their (both collective and individual) cultural backgrounds. However on the other hand, there are precise universal moral obligations which should be observed prior to legitimize the outcomes of siting processes (Peterson and Hansson 2004). Identifying rigorous and consistent ethical *principles* through which justifying the legitimacy of siting processes beyond any contextual specificity remains, therefore, a needed research effort.

Deciding about technological risks entails deciding about how, in order to (supposedly) advantage the whole, a status and feeling of uncertainty can be (concretely) imposed to what will become disadvantaged ones. A status of risk inequality among members of society is

indeed a common implication of the siting of risky facilities. Notably in the case of site-specific installations, the created *risk* inequality is largely a form of *spatial* inequality. This is why the matter of risk inequality is the point of departure of the present discussion, whose most general question is: What is *concretely* at stake during risky facilities siting processes? When taking the predominant NIMBY outlook the immediate reply would be the trade-off between the collective benefits brought by the technology under consideration and the local risks imposed to those living nearby. Weighing such collective benefits against the local safety loss is therefore what ought to be done. The assumption is that both gained benefits and safety loss can be calculated and then used as the indicators for describing the “siting problem.” Is that the case?

The question may sound banal. The benefits brought by hazardous technologies are evidently not only quantitative entities of immediate appreciation; for example, somehow and someday nuclear power production could reveal to have been the greatest contributor to climate change abatement. By its side, although perhaps less evidently, safety is more than the antonym of risk (Moller et al. 2006). In a given spatial setting, *being safe* and *valuing* such feeling of safety are determinant factors of opposition to intruding technologies (Simmons and Walker 2004). Eradicated “space values” result from the intra-subjective meanings deposited by individuals on other individuals and, through them, projected on the shared living environment. In other words a community that shares a history and therefore a feeling of safety attaches a profound value to it, and through it attaches a value to the own living environment. In short, safety is not only a quantifiable status related to the major or minor exposure to risks; it is also, when not primarily a value at the very basis of the identity of places.

To conclude the unquantifiable and intangible value-component of safety that is cemented in the collective perception of the space shared by individuals is what is concretely the object of violation when a risky technology is proposed for siting. The siting of an ultrahazardous facility puts at risk not only people and things but also the value and identity that will be attached to the place in which they live. This seems to be the underlying idea also of the analysis of Boholm regarding the main areas of investigation to be further explored for shedding new light on siting controversies (Boholm 2004). These areas are listed below:

1. Identifying appropriate procedures for planning and decision making by accounting both the tangible and intangible “values at risk” of the localities intruded by controversial installations
2. Identifying the distinctive aspects of NIMBY cases in contrast with the overall lack of acceptance of controversial technological developments which may be signaled during conflicting siting processes
3. Promoting a broader understanding of the consequences of risks in terms of the threats to the space values that a community regards as nonnegotiable, and accounting them since the early stage of siting processes

This work discusses these three areas of investigation but will contribute mainly to the development of the third one. As mentioned above, the main scope of this study is proposing a perspective on risky facilities siting based on a synergy between risk, ethical, and spatial planning theories. The contributions which will be recalled along the text are therefore both of philosophers of risks (Hansson and Peterson 2001; Peterson 2001, 2003; Peterson and Hansson 2004; Ersdal and Aven 2008; Asveld and Roeser 2009; Roeser 2010) and of ethicists of spatial planning (Moroni 1993, 1994, 1997, 2004, 2006; Golany 1995; Hare 1999). Basic notions of risk

analysis will be also provided in support of the reading (Apostolakis 1990, 2004; Amendola et al. 1992; Christou and Amendola 1998; Amendola 2001; Christou 1998).

It is worth repeating that arguing on the siting of technological risks suffers from several limitations when done at a high level of generality. Procedures for planning and decision making vary depending on different legislative, political, and cultural frameworks (Cozzani et al. 2006; Basta et al. 2008; Basta 2009). Furthermore, the type of technology under consideration and the story line of the relevant siting process in its time and space pose extremely specific challenges. This chapter opts for investigating the distinctive elements of risky facility siting processes beyond such specificities and, to do so, it starts with providing a general outline of the “siting risks” issue.

## The “Siting Risks” Issue

---

In a highly simplified manner, a risky facility siting process consists of the building of an installation meant to deliver certain benefits in a specific geographic setting (Boholm 2004). To such benefits certain risks correspond. Generally, whereas the produced benefits are broadly collective the so-called unwanted consequences are mostly locally distributed. Such consequences consist, essentially, of health and environmental effects.

Although effective, this introduction does evidently neglect some not banal aspects of the problem. That the unwanted consequences of technological risks are locally distributed, for example, is a consideration that has been put under severe discussion in past literature. The risks posed by certain ultrahazardous installations (e.g., by nuclear waste disposals or by dangerous chemicals’ factories) are “no more tight to their place of origin” (Beck 1986:1992) as their effects may overcome geographical as well as generational horizons. The consequences of the sadly known nuclear accident of Chernobyl of 1986 cannot be reduced to the immediate and long-term effects on the citizens and the environment of the surrounding village. Equally the consequences of the chemical disaster which killed thousands in Bhopal in 1984 cannot be understood as merely “local” (see Shrivastava 1992 and 1996 in particular). Notably while this writing undergoes the final editorial work, the consequences of the Fukushima nuclear disaster that followed the earthquake in Japan of March 2011 occupies the world headlines; even here it would be hard defending that such consequences are merely “tight to Fukushima.”

These considerations serve to clarify from the outset that notwithstanding the acknowledged multidimensionality of the consequences of site-specific technological risks this chapter will refer mainly to *spatially relevant* consequences (Schmidt-Thomé 2006a, b, c). The latter are those rare, unwanted, and disruptive consequences that following accidents develop into scenarios that involve the surroundings of installations. Differently than the *ordinary* impacts related to the operational standard of the given technology (like emissions and nuisance), these scenarios may be also identified as its *extraordinary* impacts (Basta 2009). In simpler terms, they are the so-called major accidents which may occur due a failure of the technology and/or of the human–technology interaction.

The spatially relevant measures of prevention of the consequences of major accident scenarios include safety distances, emergency planning, geographical information systems, and additional technical measures (in the following: ATM) such as reinforced glasses and protective fences. Generally, they include all the material and organizational measures

dislocated in a given territorial setting that are designed according to the risks posed by the technology (Basta et al. 2006) and are meant to prevent the relevant consequences. Spatial planning instruments (particularly, local land use plans) are among the governance instruments through which such measures are regulated, designed, dislocated, and maintained in time.

Nevertheless these instruments can be unprepared to regulate such measures timely and effectively. New and emerging technologies pose new risks to which preexisting risk prevention regulation and land use planning procedures are often neither ready to respond, nor flexible to adapt to. Furthermore the hydrogeological, morphological, natural, and urban features of the areas where newly designed technologies can be appropriately located must respond to specific requirements, which are sometimes as diverse and complex as the social contexts in which these technologies are to be sited. These features of the territory, both in the West and in the East of the industrialized world, keep varying also in time and generally shortening the possible distance between risks and inhabited urban areas. “Regions of risks” kept changing their configuration and concentration, but generally they kept expanding in both European and extra-European countries (Hewitt 1997). During the 1960s and 1970s, the siting of hazardous installations was evidently less constrained by the proximity of urbanized areas than nowadays.

Because of their political and legislative implications, planning instruments and procedures have rarely anticipated the urging matter of this shortening distance. Planning instruments, risk regulation, and risky technologies do not simply evolve at the same rhythm (Arcuri 2005). Rather the former two adapt to the demands of the latter and, in most cases, once the latter needs to be sited. Sadly, planning instruments did often adapt to risks also when the latter manifested themselves in the form of tragic consequences. This is why current European national frameworks for spatial planning in at-risk areas are often seen as *lessons learned from accidents*, with the meaningful example of the thoroughly revised French regulation on the siting of dangerous establishments which followed the accident of Toulouse in 2001 (Salvi et al. 2005).

Analyzing the postaccidents development of national spatial planning regulations is particularly useful in the context of this chapter. Different methodological orientations adopted in Europe offer the opportunity to reflect on the specific as well as cross-boundaries aspects of the “siting risks” issue. Before developing the discussion in this direction, it is important to keep in mind some clarifications. First of all, the technologies under considerations in this work are site-specific hazardous technologies of known risky potential. In more simple words, their operation implies the presence of substances and/or risk-factors known to be harmful for the man and the environment. As mentioned above, chemical factories, nuclear facilities, fossil fuel, and non-fossil fuel storages together with their distribution networks and CO<sub>2</sub> underground disposals are examples of such technologies. A second clarification is that their overall desirability will not be object of discussion. This chapter does only discuss the ethical implications of their siting. This choice is motivated by the spirit of generality that animates the present collection and furthermore by the fact that the desirability, acceptability, and justifiability of risky technology relates, at least in part, to the specific context of their design and operation. Discussing whether and how given risky technologies are desirable without accounting these aspects is therefore not possible at this level of generality. This chapter does rather wish to provide critical instruments which may support a better understanding of the complexity of spatial planning processes in relation to technological risks and,

to do so, only their most common features can be considered. Possible routes to be followed to deviate from generality are provided through the bibliography.

## History

---

The matter of siting risky technologies became a prominent topic of investigation for increasingly heterogeneous scholarly fields starting, in particular, from the 1970s (Ale 2005a, b). At that time the increasing presence of chemical establishments, LPG storages, and networks of fuel distribution started to conflict with the risk of accidents within, equally expanding, urban areas. Periodically revamped by the occurrence of catastrophic accidents, the matter of “siting risks” remains one of the most controversial challenges posed by technological development to spatial planning practices. Considering the parallel growth of urban areas and industrial sites that characterized the last 60 years of European history, evidently the “space available to risks” has been decreasing (Hewitt 1997) while, at the same time, the catastrophic potential of hazardous technologies kept dramatically augmenting (Beck 1986;1992).

New technologies imply new risks whose modeling and quantification is a complex exercise of sometimes debatable outcomes. However also the *ex ante* assessment of the risky scenarios associated to established technologies (like LPG storages or chemical factories) is characterized by a degree of uncertainty and subjectivity (Amendola et al. 1992; Apostolakis 2004). Stochastic uncertainty (i.e., the variability of the variables in time; e.g., of predominant winds direction) and imperfect knowledge (i.e., the lack of information) limit the possibility of predicting the development of accident scenarios with precision (Christou 1998). Several studies have also demonstrated that on the basis of the same information experts called to perform a risk analysis exercise may arrive to different, although all well-grounded, conclusions (Amendola et al. 1992). These aspects of uncertainty and subjectivity do not limit the relevance of risk analysis to risk decision making, but do surely warn against an a-critical utilization of its outcomes.

The European Directive stating the land use planning requirements regarding the siting of establishments wherein hazardous substances are present (the before mentioned Directive 96/82/EC on Dangerous Substances) is named after the accident of Seveso occurred in 1976 in the Northern region of Lombardy (Italy). The accident consisted of a massive release of dioxin and involved the population living in the nearby rural area. Besides causing a number of injuries and precautionary abortions, the accident led to a long lasting pollution of the soil and, consequently, of the food chain. The emotional and political impacts of these dramatic consequences were proportional to the unpreparedness of the population of Seveso to cope with them. Citizens were unaware of the nature of the substances produced by the establishment. Meaningfully, because of the pleasant odors released during the processing of chemicals the plant was nicknamed “the fabric of perfumes” (Arcuri 2005). Information on the effects of dioxin on edible vegetables, animals, and above all humans was delivered by the plant management to authorities and the public with dramatic delay. The evacuation of the area surrounding the plant following the accident was consequently organized with substantial delay. Despite following accidents worldwide led to worse immediate consequences in terms of human lives and environmental impacts, the Seveso accident does therefore remain the epitome of everything; the spatial and emergency planning practices must be prepared to prevent and to cope with.

It is also by reflecting on the many implications of these tragic accidents that starting from the 1990s different scholarly fields started to contribute to the technological risks literature. The risk analysis literature (Apostolakis 1990; Christou and Mattarelli 2000; Amendola 2001; Cozzani et al. 2006 among others), the legal and sociological literature (Beck 1986:1992; Renn 1992; Arcuri 2005; Renn and Graham 2005), and the geographical and spatial planning literature (Walker 1991, 1995; Boholm and Löfstedt 2004; Lesbirel and Shaw 2005) started to have, in the matter of “siting risks,” a shared area of interest. Psychologists started to contribute to the understanding of (to paraphrase one of the most known authors among them) the interpretational distance between “risk as numbers” and “risk as feelings” (Slovic 1991, 1999, 2002). Thanks also to the empirical evidence provided by these studies that ethicists could start reflecting on the moral implications of such distance (Roeser 2006, 2010), while others kept arguing on the inherently ethical nature of the analytical assessment and decision-making processes whose objects are the most diverse risk-bearing technologies (Schrader-Frechette 1991; Peterson and Hansson 2004; Ersdal and Aven 2008; Asveld and Roeser 2009; Hillerbrand 2010).

Spatial planning started to contribute to this heterogeneous research area somehow later than the listed scholarly fields. One of the possible explanations of this delayed interest from the side of planning scholars in a matter that is increasingly affecting planning processes worldwide is the historical distance between spatial planning theory, spatial planning regulation, and the relevant professional practice. But for few exceptions (see Walker 1991, 1995) before the current decade spatial planning scholars had hardly considered the “siting risks” issue as an autonomous field of investigation. The relatively recent inclusion of land use planning requirements for areas at risk within European legislations might explain this otherwise moderate attention. Notably the land use planning requirement of the mentioned Seveso Directives (the already mentioned Article 12 on the Control of Urbanization) was added to the requirements of the Directives relatively recently, especially when thinking at the events that led to its formulation. Article 12 is dated 1996, hence two decades “older” than the Seveso disaster and a significant number of disastrous accidents worldwide (see Lees 1996 for a thorough account). Several national transpositions of Article 12 are also quite recent. The Italian Decree on land use planning in areas subject to major accidents risk was issued in 2001; the new French Law on natural and technological risk prevention in 2003 (see Basta 2009). The majority of contributions of spatial planning scholars followed but have rarely anticipated the methodological challenges posed by these new regulatory requirements. It could be said that as risk regulation is a lesson learned from accidents, spatial planning derives lessons from risk regulation.

A second possible explanation of the relatively scarce contributions of spatial planning scholars to the risky facilities siting literature is the historical “ownership” of the analytical assessment that informs siting processes by the side of risk analysts. Somehow when it comes to large-scale hazardous facilities, land use planning instruments become the “recipients” of (risk) assessments solely performed by technicians. The most impacting technological risks are still too often “implanted” within local land use plans without having preventively integrated the knowledge of risks with the knowledge of the territory. The latter is intended not only as the material configuration of the living environment but also as the immaterial dynamics at the basis of its identity and future developments. The knowledge and understanding of these dynamics would be fundamental to site hazardous installations compatibly with the surroundings and with their lines of development.

Next to the historical ownership of locational assessments by the side of risk analysts, there is also the problem of the different evaluative scales of risky facilities siting and local land uses. In the majority of European spatial planning frameworks, the central or regional governments hold the competence for licensing top-tier Seveso establishments and, generally, risky facilities of “general interest” like nuclear power plants and energy infrastructures. Differently, the elaboration of land use plans is usually assigned to municipalities with all the deriving problems of coordination of consistency. There is, in essence, a problem of “who does what” and “at what scale she does so.”

The lack of coordination between the technical assessment that guides the siting of a given risky technology and the procedures for allocating land uses and granting permits of construction in the same area resulted in the many situations of proximity between risks and residents which are visible in European cities. Installing technological risks within land use plans “from the top” while dealing with the development of industrial sites separately from the development of the surrounding areas equaled to a sort of technological colonization of localities or, as defined elsewhere, of industrial encroachment (Simmons and Walker 2004). Several spatial planning scholars have captured the unpreparedness of spatial planning instruments and of the professional practice behind them in preventing a process of siting from becoming, in time, a sort of uncontrolled industrial metastasis. In a somehow paradigmatic case study, Walker and Simmons (2004) documented how in the course of five decades a small industrial area in a village in the South of Great Britain kept expanding up to 17 ha. At the same time residential districts expanded, if not proportionally, equally considerably. The proximity between the boundary fences of the industrial site and nearby houses augmented especially during the 1970s, with the current result that the fences have little or no separation from gardens and footpaths. What this situation exemplifies is that the land use planning in the area had not foreseen the situation of conflict which would have resulted from the mutual expansion of the industrial area and of the residential districts; most probably because of the lack of coordination between the two respective licensing processes. On the one side the industrial management was allowed to acquire new land and expanding the works, while on the other side developers and privates were allowed to acquire new land and constructing residences. The current situation of proximity is therefore difficulty resolvable as it consists of the opposition between parties who have acquired, in time, equally legitimate property rights.

As any other case of industrial encroachment, the story line of this case is highly contextual. However, it is paradigmatic of the historical lack of a methodological and procedural bridge between risk prevention and land use planning. Furthermore, it is paradigmatic of the impacts on the self-identity of communities that end up with cohabiting with industrial establishments, and the associated risks, in an increasingly intruding and stigmatizing way. Hazardous industries may progressively stigmatize the perception of certain areas to the point that what in the collective memory was recalled as a (safe, relaxing, unpolluted) rural area does quickly become an (unsafe, threatening, polluted) industrial one. Somehow, different values and attributes start to be associated to the living environment once hazardous installations become a substantial part of it. How to incorporate the values that are attached to the living environment before its potential installation (e.g., the valued feeling of safety) when it comes to decide about their best location is therefore a crucial point of attention for the planning research in the field.

As anticipated, in order to develop this discussion, the main areas of investigation identified by Boholm (2004) helped in organizing this chapter in the following sections.

## Current Research

### Finding a Place to Risks: The Example of the European Regulatory Framework on Dangerous Substances

In order to discuss the *methods and procedures* for spatial planning in relation to site-specific hazardous establishments, this section refers to a European chemical safety regulation, that is, the already introduced Seveso Directives on Dangerous Substances (96/82/EC and following amendment). Article 12 of the Directive states the Control of Urbanization requirement. An overview of selected national implementations of the requirement will help to identify some specific aspects of national planning methods and procedures while highlighting the cross-boundaries aspects of the “siting risks” issue.

Article 12 calls Member States to “ensure that the objectives of preventing major accidents and limiting the consequences of such accidents are taken into account in their land-use policies and/or other relevant policies.” A major accident consists of the release, explosion, or fire involving dangerous substances whose impacts may extend outside the boundaries of establishments. The massive dioxin release in Seveso (Italy, 1976), the disastrous methyl-isocyanate release in Bhopal (India, 1984), the fireworks explosion in Enschede (The Netherlands, 2000), and the ammonia explosion in Toulouse (France, 2001) are all examples of major accidents. Considering their risks while elaborating land use planning instruments consists, essentially, of modeling possible accident scenarios and “overlapping” their visualization with the land use map of the area. To model major accident scenarios, risk analysts consider a great number of variables, including the stock of dangerous substances, the safety system in place, and the predominant meteorological conditions in the area. Visualizing the spatial extension of accidents on a geographical basis does therefore serve to identify whether the corresponding land uses (residential, service, or transportation routes) would lead to the involvement of residents and environmental goods. The following step is establishing the spatial measures which may prevent it. The most important spatial measures are safety distances and emergency plans. According to the Seveso Directives, safety distances in particular should restrict land uses and being maintained in the long term.

Following the entry into force of the Seveso Directives, in 2004 the Major Accidents Hazards Bureau of the Joint Research Centre of the European Commission launched the “Land Use Planning Including MAHB and NEDIES” investigation. The focus of the investigation was on the different approaches adopted by Member States for implementing Article 12 within their land use policies and instruments. The scope of the investigation was providing an up-to-date overview of national implementations of the requirement and supporting the elaboration of a European implementation guidance. The latter was adopted by the European Commission in 2006 (Christou et al. 2006).

Next to the implementation guidance, the MAHB investigation led to the elaboration of a second supporting instrument, published in the form of a JRC Technical Report, named *Roadmaps* (Basta et al. 2008). Based on a questionnaire survey, literature review, and direct interviews with the members of the European Working Group on Land Use Planning (EWGLUP), this JRC technical report collects the detailed description of the different methods and procedures for land use planning in Seveso areas developed by a selected group of Member States. The Dutch, British, German, French, and Italian implementations of Article 12 are here analyzed together with the procedures in place for their implementation within land use

planning instruments. Indications for interpreting the differences among these five member states' *roadmaps* from Article 12 to local land use plans are provided together with general recommendations to member states on how implementing the Article in accordance with the European regulatory framework.

These different European national implementations of Article 12 are particularly interesting in the context of this chapter as they allow to highlight fundamental differences in spatial planning frameworks across Europe and the way different risk prevention policies interrelate with them. As documented in a case study that was performed during the elaboration of the implementation Guidance, the "methodological bridge" between the two national regulatory frameworks (that are the spatial planning and the major accidents risk prevention frameworks) leads to remarkably different land use planning interventions (Cozzani et al. 2006). Understanding the underlying factors of such differences while identifying what does not vary across methodologies, boundaries, and cultures has therefore been the focus of this chapter. Before introducing some findings such differences are briefly described.

From a methodological perspective, major accidents risk assessment in relation to land use planning can be distinguished in probabilistic and deterministic approaches (Christou 1998; Cozzani et al. 2006; Basta et al. 2008; Basta 2009). The assumptions at the basis of the two approaches differ in relation, particularly, to the decisional relevance of the expected probabilities of accidents. The probabilistic approach considers the calculated probabilities of accident scenarios as an explicit decisional criterion of the assessment of safety distances from dangerous establishments and of the compatible land use destinations in their surroundings. Differently, the deterministic approach considers as explicit decisional criterion only the expected consequences of accident scenarios.

This difference is fundamental as, in principle, it implies that a very rare accident scenario (say with less than  $10^{-8}$  event/year probability of occurrence) with disastrous consequences (immediate death of stationary individuals within, say, 500 m of distance from the source of the accident) may be discarded when land use planning instruments are to be designed. On the basis of its extreme rarity, the probabilistic planner who has to "draw on the map" the safety distance which shall be restricted from construction may simply ignore it. Differently, the deterministic planner would not regard the rarity of the event as an explicit decisional element: Only its consequences would be considered as such; therefore, the distance within which the consequence of "immediate death" would occur, regardless of its expected probability, would become the reference for drawing safety distances on the map. The rationale of the deterministic approach is that an even poorly credible catastrophic event (that in technical jargon does usually correspond to the *worst-case scenario*; see Lees 1996) would be considered as the *reference-scenario* on the basis of which safety measures could be established. It is important to point out that the most adopted form of determinism in major accidents risk prevention is selecting, in place of the worst scenario, the worst *credible* one (accounting, in so doing, probabilities implicitly rather than explicitly). Nevertheless, the rationale of the probabilistic and deterministic approaches remains sharply different and as demonstrated leads to substantially different land use planning interventions (Cozzani et al. 2006).

The United Kingdom and the Netherlands have adopted and largely contributed to the development of the probabilistic orientation (Ale 2005a, b; Bottleberghs 2000; Jongejan 2008). Germany has historically maintained a deterministic orientation. France and Italy have recently reviewed and formulated their legislations respectively and have both opted for a mixed method. The latter method decays from the deterministic orientation as it privileges the

consideration of the severity of the consequences of reference-scenarios while accounting expected probabilities as a mitigating factor (for a detailed overview of these national methods see Basta et al. 2008).

A second significant difference among the analyzed national approaches to land use planning in Seveso areas is the target versus legally binding risk tolerability criteria which are indicated in legislations. Adhering to its common law tradition, the United Kingdom does not provide fixed criteria: The as-low-as-reasonably possible (in the following: ALARP) principle underlies the risk prevention policy of the country (Ale 2005a) and no legally binding, but only target criteria, are indicated in the relevant legislation. With the exception of the Dutch societal risk tolerability criteria, generally European national legislations provide a legally binding risk tolerability or “consequences acceptability” threshold that has to be respected when establishing the land use planning measures in the vicinity of hazardous establishments.

• *Table 11.1* summarizes the criteria used for major accidents risk assessment in relation to land use planning in the examined countries.

■ **Table 11.1**

Risk in land use planning: a summary of policy orientations in selected European countries

	MA risk prevention policy orientation	Criteria for risk assessment in land use planning	Status of the criteria
France	Traditionally deterministic; recently reviewed, it accounts severity levels and probability classes for determining safety distances	Individual risk and “severity” levels on the basis of the number of expected fatalities and damages	Legally binding
Italy	Semiquantitative: a deterministic assessment of accident scenarios and effects is followed by the consideration of probability classes as mitigating factor for defining safety distances	Individual risk and environmental risk	Legally binding
Germany	Traditionally deterministic	Individual risk	Legally binding
The Netherlands	Strictly quantitative: a full QRA is required in all cases risky establishments have to be installed and operated	Individual and societal risk	Individually binding for individual risk, target criteria for societal risk
Great Britain	Based on the application of the ALARP principle, the evaluation of safety distances is risk oriented in case of emissions and consequence oriented in the case of thermal radiation and explosions	Individual and societal risk	Target criteria: the advice of the safety authorities is not legally binding for planning agencies

The Table provides a general summary of the approaches adopted by the selected European countries for regulating land uses in the vicinity of the hazardous establishments falling under the requirements of the Seveso Directives. How to explain the adoption of different methods by the side of often neighbors Countries? Which are the most relevant spatial implications? Both questions are discussed in the following section.

## Deciding on Chances or Deciding on Consequences? European Answers and Their Ethical Basis

As discussed above, the first and most important difference between the approaches adopted in selected European countries for land planning in at-risk areas consists of the adoption of deterministic versus probabilistic approaches. The spatially relevant implications of the two approaches could be summarized as more conservative and sensitive to hazardous substances stock in the former case and less conservative and more focused on risk abatement at source in the latter case. Deterministic approaches do usually consider, as reference-scenario, worst-, or worst-credible scenarios. By doing so they lead to opt for larger safety distances, for decreasing the stocks of substances in order to shorten them or, at least theoretically, for (re)locating the establishment elsewhere in case both measures reveal insufficient to site the establishment compatibly with the surrounding land uses. Probabilistic approaches instead rely on quantitative risk assessment (QRA) for modeling the reference-scenarios which may *credibly* occur and select the reference-scenarios to be used for planning purposes on the basis of a given risk tolerability threshold. In the Netherlands for example, the risk of  $10^{-6}$  events/year probabilities for a stationary individual of dying due to the involvement in a major accident is the legally adopted risk acceptability threshold (Bottleberghs 2000; Ale 2005a, b). Above this threshold, accident scenarios are therefore not considered for planning purposes.

Even though the described divergences between the two approaches are more theoretical than practical, the range of considered accident scenarios does concretely differ. Whereas the probabilistic approach leads to restrict the set of reference-scenarios to those whose occurrence is credible from both a magnitude and probabilistic viewpoints, the deterministic approach considers the whole range of possible events, including the worst. As worst-scenarios are likely also the rarest and have the most extended area of impact it follows that the spatial measures in their relation include, among other measures, long safety distances and large restricted zones.

There are several possible explanations of this fundamental difference between the two described approaches. One relates to the national environmental and spatial frameworks in which they are adopted. In countries like Germany, wherein the Federal Constitution defines the precautionary principle as a fundamental environmental principle, justifying the adoption of a probabilistic orientation in risk prevention may be regarded as inconsistent (Boehmer-Christanses 1994, reported by Adams 2002). Differently in the Netherlands, “the government has committed to the concept of risk rather than to the false promise of safety” (Jongejan 2008). Here the combination of land scarcity and high population density represents an insuperable limitation to the adoption of a deterministic orientation: Restricting large areas around hazardous establishments from construction would conflict with the need of other primary land uses.

Next to policy and spatial issues also economic considerations matter. Current situations of proximity between hazardous establishments and urbanized areas consist of the opposition

between parties who have acquired, in time, legitimate property rights. Any measure of relocation and/or land use restriction would hence require a large investment of both public and private resources. For new siting processes, the problem is somehow similar: Imposing large clear areas around Seveso establishments may lead to a loss of constructible land plots and hence conflicting with other primary residential and infrastructural needs.

These considerations are, however, more of pragmatic than of theoretical nature. From a theoretical viewpoint, the rationale of the two approaches could be rather seen as the result of fundamentally different ethical assumptions. Opting or not opting for considering a threshold of probability of occurrence of unwanted scenarios as an explicit decisional element for selecting the reference-scenarios to be used for planning purposes relates to relying or not relying on quantitative risk assessments of known uncertainty. In other words, it relates to considering or not considering the whole range of consequences regardless of the investment needed to limit their possible, although extremely rare, occurrence. In conclusion, it relates to “taking the risk” of relying on quantitative risk analysis. As mentioned above, imperfect knowledge and the unpredictability of contour-conditions in time and space make the modeling of accident scenarios a necessary but not exclusive criterion on the basis of which spatially relevant risk prevention measures were established. Note that the engineering jargon in this field is full of qualitative attributes: the “credible,” “worst,” “poorly credible” labels stress the qualitative component of the *ex ante* assessments whose objects are events whose likelihood and development are not fully predictable.

From a theoretical perspective, it could therefore be said that deterministic and probabilistic approaches differ on the ground of their different attitude toward uncertainty. Determinism is bent toward the maximum avoidance of it. This could be interpreted as the prioritization of a moral principle, that is, the principle according to which what the planner ought to do in front of the *possibility* of unacceptable consequences cannot be mitigated because of their extreme rarity. What ought to be done is opting for their avoidance *tout court*. Probabilism instead moves from the assumption that the planner ought to consider the *probability* of unacceptable consequences in light of the sacrifice needed for their prevention. What ought to be done is therefore preventing (only) the set of credible consequences by considering the (calculated) rarity of events as an (explicit) decisional criterion.

The former assumption, that is that the whole set of *possible* scenarios must be object of prevention, is based on a categorical imperative. The latter assumption, which differently sees only the *credible* scenarios as the due object of prevention, underlies a balanced consideration of the relation between risks and the resources needed to prevent it. In other words, it considers the trade-offs between risks and benefits. On the basis of these remarks, it is proposed that the deterministic and probabilistic approaches for land use planning in at-risk areas may be linked to deontological and utilitarian theories respectively.

Several authors have investigated the implications of different ethical theories to risk decision making and only a dissatisfying selection can be cited here (Schrader-Frechette 1991; Peterson 2001, 2003; Peterson and Hansson 2004; Ersdal and Aven 2008). Ersdal and Aven (2008) for example give an account of the moral theories which can be retraced underneath the surface of the risk assessment criteria which are most commonly adopted by legislation and by professional practices. In their account, utilitarianism is defined as a theory of both the good and the right as it identifies both a nonmoral value (i.e., utility) and which principle shall justify the right actions (i.e., its maximization). Deontology by contrast is a theory solely concerned with establishing the principles following which actions would result

good or bad. Its most known formulation, that is the Kantian formulation, is based on the concept of categorical imperative, which is the universal normative guidance through which judging the “goodness” and “badness” of actions. A categorical imperative is universal in the sense that it applies to anyone, anywhere, and anytime: No breaches are possible under any circumstance. An interesting interpretation of deontology in contrast to utilitarianism is that categorical imperatives are in fact not established on the basis of what individuals ought to pursue and realize, but solely on what individuals ought to honor (Pettit 1991). Therefore, while in utilitarianism the goodness of badness of actions relates to their capacity of *realizing* individual and collective utility, in deontology the same statuses relate to their capacity of *honoring* given moral imperatives.

As known utilitarianism is based on consequentialism in the sense that what individuals ought to do among, for example, possible alternatives relates to their consequences. Deontology instead rejects the idea that a moral imperative (say, “individuals ought not to pose risks to others”) can be overruled because of the consequences the breach of the imperative could imply (e.g., individuals may pose risks to others to achieve a greater collective utility). In a strictly Kantian perspective this is not solely a matter of morality, but rather of rationality: Breaching the categorical imperative opens de facto to a rational contradiction. If one assumes, categorically and imperatively, that killing others is unjust but that “now and then” it should be possible because of the greater positive consequences, than one should also accept the unbearable contradiction that “killing someone is just.” This is an unacceptable contradiction from a rigorously Kantian deontological perspective.

When applied to the two approaches to land use planning in at-risk areas that are described above, the utilitarian and deontological theories reveal to constitute the assumptions of the probabilistic and deterministic orientations respectively. The probabilistic planner grounds her decision on the consideration of its consequences, which are, to avoid any confusion, not the consequences of the events under consideration but solely the consequences of considering or not considering extremely low probabilities of events as an explicit decisional criterion. One of these consequences, and perhaps in utilitarian terms the most important, is the high cost of referring to the worst and most unlike scenarios each time safety measures have to be established. In the specific case of hazardous facilities, siting this is particularly relevant as spatially relevant safety measures are extremely costly in terms of construction capacity loss. Is this capacity loss justifiable when considering the extreme rarity, if not poor credibility, of worst-case scenarios? The latter result from the combination of all worst conditions, for example the maximum amount of substances possibly present within the establishment, the worst meteorological conditions, the worst possible chain of technological failures, the highest possible number of contemporarily exposed individuals, etcetera. From a probabilistic point of view, the contemporary combination of all these conditions may indeed be so extremely rare to be unrealistic. At the same time, their disastrous nature would result in unduly high costs of prevention. These considerations lead the probabilistic planner to opt for a more realistic appraisal of the risks under consideration and thus to opt for credible scenarios in order to establish the safety distances to be maintained from establishments.

One of the moral implications of this line of reasoning is that by acting according to considerations of “credibility” and “utility” the planner, in case of occurrence of the “incredible” scenarios she has not considered, discharges the responsibility of their consequences. This is explicitly stated in a milestone of the British risk prevention policy that reads “[...] a computation must be made by the owner in which the quantum of risk is placed on one

scale and the sacrifice involved in the measures necessary for averting the risk (whether in money, time or trouble) is placed in the other, and that, if it be shown that there is a gross disproportion between them – the risk being insignificant in relation to the sacrifice – the defendants discharge the onus on them.” (British Court of Appeal in its judgment in Edwards vs National Coal Board, 1949, 1 All ER 743). In this sentence, arguably the “gross disproportion” between an “insignificant” risk in relation to the “sacrifice” needed to prevent its consequences has a clear normative relevance. According to it the planner would *not* be responsible for *not* taking action in all cases in which she can demonstrate the existence of such gross disproportion. This equals defending that spending enormous amounts of resources for preventing events whose occurrence is poorly credible is not what the planner ought to do, to the point that she discharges the onus of their consequences even in the case of their occurrence. Following this rationale, one could conclude that the level of individual responsibility relates to the level of (in)significance of risks.

Yet insignificant risks do not imply insignificant consequences. To the contrary, when it comes to site-specific hazardous technologies very low probabilities of occurrence of accidents are usually coupled with the most disastrous effects. In contrast with the probabilistic planner, the deterministic one would therefore act on the basis of a categorical imperative, that is, that individuals ought not to pose the risk of unacceptable consequences to others regardless of their probability. On the basis of this principle she would not embark on a risk/benefits calculation of given alternatives, but would only honor the principle of not allowing some to risk of causing unwanted consequences to others. To her the *possibility* of such consequences is what truly matters; their risk is not relevant to her decisions. Thus, the worst possible consequences would become her reference-scenario and she would establish the spatial measures preventing third parties from suffering from them. Restricting the land uses around the installation up to the distance which would prevent inhabitants from being involved in the consequences of the worst possible accidents would therefore be her most rational, and only morally justifiable action. The even smallest probability of dying due to a poorly credible accident would be eligible to become her reference-scenario. Any other decision which would contrast with the categorical imperative would lead to a contradiction that the deterministic planner would find irrational beside immoral: If she would accept that extremely low probabilities of occurrence of accidents may be considered as mitigating factors when dimensioning and establishing safety measures, she would have to accept that “individuals may pose risk to others.”

As the difference between the two described lines of reasoning is primarily due to a different attitude toward the relevance of probabilistic judgments to decision making and therefore, implicitly, toward the component of uncertainty in risk analysis, the probabilistic approach is defined as uncertainty acceptant whereas the deterministic one is defined as uncertainty avoidant. Both are grounded on rational assumptions and, similarly, lead to some rational distortions. The most evident distortions are that the probabilistic planner, in order to justify her action, needs to establish above which level risk becomes negligible and can hence be ignored. By her side, the deterministic planner will have to deal with the matter of infinitude of the conditions which may “worsen the worst.”

The former issue, that is the problem of setting a risk tolerability threshold under which decisions are “the right” decision, has been already debated in moral philosophy (see Peterson 2001, 2003 in particular). The de minimis risk, or “so low to be negligible” risk, is a controversial concept. The very attribute of “negligibility” obscures the component of

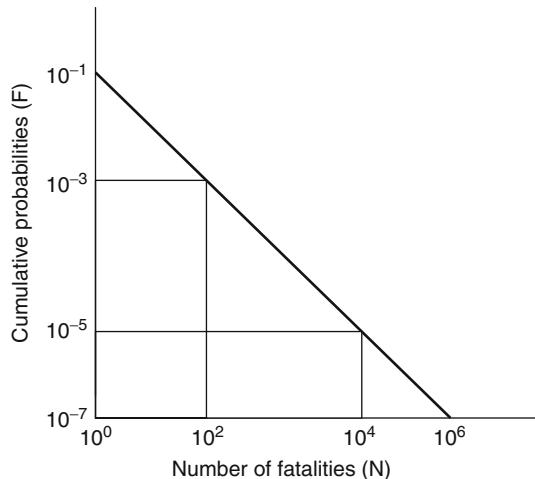


Fig. 11.1  
Fictitious FN criterion (Adapted from Jongejan 2008)

consequences that can be, by contrast, extremely disastrous. Furthermore, setting a quantitative limit under which establishing the de minimis risk is confronted with the problem of comparability: Mathematically speaking, two risks may be equivalent while having two different expected probabilities. This is illustrated in the following (fictitious) FN graph (Fig. 11.1).

The graph illustrates a fictitious FN criterion (i.e., the diagonal bold line), that is the acceptability threshold for a number of individuals  $N$  to die due to the involvement in a major accident with cumulative frequency  $F$  (for reasons of simplicity the difference between cumulative frequencies and cumulative probabilities is ignored here). As shown in the figure the rationale of the FN criterion is that the more probable the event, the lesser the number of acceptable fatalities. Any area below the FN criterion is equivalent, as the expected fatalities are the same for any product  $F * N$ .

Nevertheless, it is hard to defend that an extremely rare event causing the sudden death of 10,000 people is both politically and socially “as acceptable as” more probable events causing 100 deaths. Although statistically the cumulative consequences of the two events are the same, they cannot be considered equivalent in terms of gravity and postaccident response. The judgment that the sudden although rarer loss of thousands of lives is “worse” than the more frequent loss of tens of them is, in conclusion, a consideration that does not rest from mathematics. Rather, it follows a fundamental moral intuition, precisely the intuition that the capacity of society of functioning properly from the political, psychological, and economical points of view would be much more compromised in the former than in the latter case. The popular objection that individuals should be “rational” and thus touched by the occurrence of technological disasters as much as they are by the daily loss of lives in traffic misses a fundamental point. The “indicator” of the moral and emotional responses to the consequences of unwanted events cannot be reduced to the same indicators used by analysts to describe them. Paraphrasing, feelings and figures do not necessarily go together. The sense of devastation and distrust caused by major accidents in which individuals are involuntarily

involved, and that reveal the capacity of certain technologies to simultaneously kill thousands of people, cannot be reduced to the number of deaths. Relying on “risky numbers” for establishing what is negligible and what is not is limited by the fact that “negligibility,” when related to mono-criterion probabilistic calculations, is *per se* an attribute that cannot be captured solely through its mathematical representation.

Notwithstanding the clear limitations of the probabilistic reasoning also the deterministic one leads to several controversial implications. Brought to its extreme consequences the principle according to which referring to the worst-scenario is the only morally acceptable choice (and hence the proper rational guidance for decisions about risk) implies considering the simultaneous occurrence of all worst possible conditions. But designing technologies and organizing society in such a way to prevent the scenarios deriving from their simultaneous combination would orientate the allocation of resources toward the achievement of a virtually impossible “0 risk” status. The variables which may “worsen the worst” are innumerable: Technologies pose risks not only intrinsically, but also extrinsically. Technologies may not be hazardous *per se* and yet the human–technology interaction being so. A moral orientation bent toward the prevention of all possible unwanted consequences of (the interaction with) technologies would therefore end up with restricting not only hazardous technological developments, but all possible dangerous interactions with harmless ones.

It is beyond the scope and capacity of this chapter to provide a thorough account of how the reality created by a deterministic and a probabilistic planner would look like. However, it can be concluded that the former reality may be not less “risky” than the one resulting from a balanced consideration of the probabilities and consequences of the events humans want to prevent. This part of the chapter highlighted that opting for one of the two orientations moves from precise moral assumptions. This is of great relevance to spatial planning processes as it suggests that the underlying principles guiding land use planning evaluations in relation to technological risks are indeed of ethical and not solely of policy, technical, or economic nature. The suggestion of Boholm (2004) regarding the need of more participated spatial planning processes in relation to technological risks could be therefore extended to the involvement of stakeholders in the *choice of the approach to* spatial planning evaluations before in the *discussion about* their outcomes.

As shortly discussed in this section different ethical assumptions regarding the criteria to be used for planning purposes in the surrounding of Seveso facilities give form to different methodological approaches and, arguably, lead to different planning interventions. Such ethical assumptions are at the core of their moral justifiability and should therefore constitute the point of departure of any negotiated and informed decision on the “where” of hazardous technologies. These aspects of the problem will be further discussed in the conclusive section. In the following, this chapter discusses the equally crucial matter of framing siting controversies correctly, and does so by referring to the recent case of proposed siting of a CO<sub>2</sub> storage in the earth of Europe.

## Framing Siting Controversies: The Case of the CO<sub>2</sub> Storage in Barendrecht, The Netherlands

In the course of 2009, the municipality of Barendrecht (the Netherlands) engaged an open conflict with the Dutch government in relation to the proposed siting of a CO<sub>2</sub> underground

disposal. The pilot-project should lead to the installation of this recently designed technology within the next few years and was proposed by a known international oil corporation. The technology promises to contribute to abating carbon dioxide emissions as it consists of capturing the latter “at source” and storing them into exploited gas fields. The Dutch Applied Research Institute (in the following: TNO) assessed the most appropriate location for siting the storage within the national territory and considered 12 possible locations, precisely seven offshore and five inland locations (Breunese and Remmerts 2009). Among the 12 candidate sites, the one in the inland municipality of Barendrecht is also the more densely populated. Notwithstanding this the geological characteristics of the gas field underneath its surface and the technological requirements of the disposal (among which the limited possible length of the pipeline connecting refineries to the underground storage), according to the Dutch government made of the site in Barendrecht the most suitable candidate for proceeding with the pilot-project. The alleged general motivation is that “*capture and storage of CO<sub>2</sub> is a necessary transition technology to help cut carbon emissions*” (Reuters, November 18, 2009).

The story line of the Barendrecht case recalls a typical facility siting conflict. The opposition to the pilot-project saw local versus national actors and, implicitly, local versus national interests. The “owner” of the decision-making process is the central government. Consistently with a policy objective and in concert with a private actor the government delegated the feasibility study to a (supposedly) neutral technical advisory body. The latter assessed a number of candidate sites on the basis of several criteria, among which the geological characteristics of available gas fields, the costs of each alternative, and the technological requirements to be met. Among the candidate sites, the chosen area of Barendrecht was the most densely inhabited; nevertheless, the final assessment report hangs in favor of siting the disposal in that area. In response the local community manifested against the installation, and a lively debate started to occupy the media to the point of becoming of national resonance.

One of the predominant outlooks on these types of controversies favors their interpretation as a not-in-my-backyard (NIMBY) cases. In effect, the Barendrecht example could be simply understood as a NIMBY situation, consisting of the rejection of the controversial (although collectively beneficial) installation because of its negative impacts on the locality. Such impacts are not only future and potential, but considering the loss of property values and the stigmatization of the area also immediate and concrete. However not all siting conflicts relate to the rejection of a technology that is solely “locally” perceived as undesirable and unjustly imposed. Whereas “the problem is frequently constructed as meeting some national need whilst ensuring justice for local communities” (Owens 2004), siting conflicts are often to be understood as general oppositions to the promotion of given technological developments. Such oppositions are given voice locally but mirror a general societal rejection.

When approaching the matter of framing siting controversies correctly it is essential, for the planner, to identify the distinctive elements of NIMBY versus non-NIMBY situations. Spatial planning processes have a primary responsibility in preventing, and possibly solving, NIMBY situations, but become dangerously instrumental in all cases in which the object of controversy is not the “where” but the very “if” of the technology. As argued by Boholm (2004) risk proponents tend to simplify the debate by assuming that a dichotomy between a generalized greater good and a local self-interest exists. The further assumption is that locals have narrow-minded perceptions of the risk at issue. Phenomena of “risk panics” (Sustein 2005) would be therefore emphasized as irrational, selfish, and irrelevant to the “just” policy decision; notably, in the cases of site-specific risks, the latter is a precise *spatial* decision. Because of their public

nature by following this rationale, spatial planning processes become merely instrumental: By polarizing public and needed versus private and selfish interests, the public actor gains the unspoken right of forcing decisions through the mechanisms of spatial planning procedures. This, even when the opposition manifested at local level, mirrors a general lack of consensus regarding the proposed technology. How to overcome this dangerously instrumental model?

A possible reply to these questions is offered by going back to the controversy undergone between the Municipality of Barendrecht and the Dutch government. A recent study demonstrates that the quantity, quality, and possibility of accessing the information about the CO<sub>2</sub> disposal and the plans of its installation is perceived by the majority of the citizens of Barendrecht (precisely by the 66% of the interviewed) as informative and sufficient (see Daamen et al. 2010). This should cautiously lead to conclude that the information regarding the facility was provided timely and sufficiently and that Dutch citizens have had the possibility to develop informed opinions about it. Nevertheless it cannot be ignored that 44% of interviewed had a different opinion and that the debate around the Barendrecht case extended to the national arena, leading several political and social groups to take a firm position of rejection concerning the possibility of adopting the technology at national scale. Several Dutch municipalities anticipated that they would have refused to host a CO<sub>2</sub> disposal should a site within their territory having been chosen for it in the future.

What these facts suggest is that the conflict regarding the CO<sub>2</sub> disposal in Barendrecht extended from the actual facility to the overall technology: In other words, the conflict has lost its *spatial dimension*. The spatial dimension of hazardous facilities siting processes is therefore what may help in distinguishing NIMBY cases from a more general rejection of a given technology. In more simple words, it is the indicator of whether the public debate gravitates around the “where” or the “if” of the installation. Even more importantly, the spatial dimension of facility siting controversies is the only dimension that is fully relevant to spatial planning processes. This dimension will be called, in the following, the *hic* dimension.

A complementary dimension to be considered when identifying the nature of siting controversies is the temporal dimension. Risks are posed “here and now” and have, therefore, precise *hic et nunc* dimensions. However, their prevention implies a non-dynamic continuous process whose scope is guaranteeing a status of safety in time (Weick 1987). Risks, together with its counterpart safety, are “dynamic statuses” that evolve in time according to technological and social standards together with their contextual political conditions. The acceptable risks of today may become the unacceptable risks of tomorrow and vice versa. Nuclear technology offers a good example of the twofold temporal dimension of technological risks; whereas, controversies related to the siting of nuclear installations relate to the “nunc” dimension, that is the potential future “now” of accidents, the debate around nuclear waste disposal focuses on the potential intergenerational impacts of radioactive waste (see Taebi 2010 and Taebi and Kadak 2010 among others). Similarly, the debate around CO<sub>2</sub> disposals gravitates around their immediate impacts on localities but also around the intergenerational effects which may follow the release of CO<sub>2</sub> in the decades to come. More generally, it gravitates around the overall sustainability of a transition technology that rather than cutting emissions at source stores them underground in the equally “finite” resources of exploited gas fields.

To conclude, while the “here and now” of technological risks relate to given spatial and temporal dimensions and do therefore constitute the focus of siting controversies, the matters of their intergenerational impacts and (un)sustainability relate to indefinite spatial and temporal implications. These implications challenge the overall desirability of technologies

and their discussion is not a spatially relevant, but obviously a politically relevant discussion. These remarks suggest that local oppositions to the siting of controversial installations may lose identifiable spatial and temporal coordinates and, by so doing, voicing a general social opposition to their large-scale adoption. Identifying whether the siting conflict has precise *hic et nunc* dimensions is therefore essential for establishing the proper framework through which it can be solved.

It could be argued that the inherent weakness of this line of reasoning is that only the technologies that are not object of “indefinite” spatial and temporal rejection by the side of the public could be object of legitimate siting processes. This is not, however, what this chapter wishes to conclude. Having defined the *hic et nunc* dimensions of hazardous facility siting processes as possible indicators of NIMBY versus non-NIMBY cases is not the remedy against the inevitably controversial nature of these types of decisional processes. As a matter of fact technological developments worldwide do and will involve the installation of often necessary, although disproportionately impacting technologies (Owens 2004). Differences in purposes, values, and perceptions will continue to affect the process of their siting and lead to contentious situations.

However the arguments above invite to reflect on the fact that spatial planning processes, together with their regulatory and legal implications, are not necessarily the framework through which finding the right answer; they are rather the platform wherein actors can collaborate in formulating the right question. As a matter of fact society “meets” ultrahazardous technologies at the moment of their siting, and it is only then given the opportunity to acquire knowledge on and voicing its perceptions of it. Identifying whether the definite *hic et nunc* dimensions of the siting of the installation are the real object of discussion is therefore the precondition for formulating the right question, that is whether the situation is truly a “not-in-my-backyard” situation or not.

To conclude, approaching all siting controversies as NIMBY situations is analytically incorrect besides dangerously instrumental, as it leads decision makers to focus on sharpening the legal and procedural instruments which may force decisions through the mechanisms of spatial planning processes. The following section discusses how spatial planning frameworks can support those decisional processes whose object is the sustainable, morally justifiable, and hence acceptable siting of technological risks in their space and time.

## Safety as a Spatial Value: Siting Risks as a Matter of Distributive Justice

Living in, moving to, and experiencing different places contribute to shape individual identities. The living environment as experienced by individuals throughout their lives determines feelings of affection, discomfort, peacefulness, gratification, and frustration, leading to sentiments of belonging and attachment as well as extraneity. The relation between space (in its broad physical, social, and political connotation) and identity (both individual and collective) has therefore received great attention by the side of geographical, anthropological, sociological, policy, and spatial planning scholars (Keith and Pile 1993; Hague and Jenkins 2005; Taibakhsh 2001; Brown and Raymond 2007). Not surprisingly it received increasing attention also by the side of those scholars investigating the relation between space and technological risk (Boholm and Löfstedt 2004; Lesbirel and Shaw 2005) and by

philosophers investigating how spatial and architectural design determines individuals' experiences of reality (Vermaas et al. 2008).

Local identities and the values that are collectively attached to the living environment are crucial aspects of the experience of disruption brought by a new technological installation. Walker and Simmons (2004) made a step forward in describing the relevance of such aspects to spatial planning by demonstrating that the stigmatization of a place caused by a large industrial site may affect the own sense of identity also in relation to *others'* perceptions. The slow and inexorable expansion of the industrial site of their study affected the "space values" which were at the basis of the sense of comfort in and belonging to the area before the plant was installed. People felt stigmatized because of it and sometimes manifested embarrassment toward strangers in admitting they were living in the town. Aspects like the easy access to the countryside and the quietness of the streets were individually and collectively regarded as clear values by nearby residents; but when these values started to be threatened by the emission of fumes and the risk of accidents the very identity of the place and of the self-perception of individuals were affected accordingly.

The point of discussion of this section is whether *safety* may be also regarded as a *space value* and whether its relevance to the preservation of what individuals regard as nonnegotiable values of their living environment makes of it also a *right* to be guaranteed. How this twofold connotation of safety (i.e., the value and right connotation) would affect, and eventually benefit, hazardous facilities siting processes will be then explored.

To develop this discussion, I will briefly refer to the already mentioned work of Ulrich Beck on the *Risk Society* (1986:1992) and more extensively to the work of Moroni on the ethics of land use (Moroni 1994, 1997, 2004 and 2006). In the following, I will also refer to the work of Peterson and Hansson on the residual moral obligations to be observed for justifying the opposition between risk runners and risk posers during siting processes (Peterson and Hansson 2004) and finally to the known Theory of Justice of Rawls (1971).

Beck wrote his academic bestseller 25 years ago, hence at a time in which the debate around technological risks was ignited by the most disastrous accidents in history. The most broadcasted among them (like the already mentioned accidents of Chernobyl and Bhopal) occurred in the context of developing economies. The well-known Beckian theory that certain technological risks are as unequally distributed in society as wealth and that the contemporary risk distribution is therefore "unjust" hence found poor, if any, resistance (Beck 1986:1992). Site-specific risks are evidently among these unequally distributed threats.

The relevance of these arguments to the present discussion will become more evident after the introduction of the work of Moroni (1994, 1997, 2004, and 2006). In his work on the ethics of land use (1997), the author investigates how different ethical theories have historically influenced and keep being at the basis of spatial planning theories. For example, Moroni discusses how utilitarianism was evidently mirrored by the practice of "zoning" which dominated the land use planning approach of the second half of the past century at European level, or how libertarian theories would convert the predominantly detailed-oriented planning interventions to more open forms of structure-oriented interventions.

Among the ethical theories analyzed in this work, the most relevant to the present discussion is the Rawlsian theory of justice, whose alleged purpose is overcoming the distortions of utilitarianism in favor of a newly formulated deontological theory (Rawls 1971).

As mentioned in a previous section, the utilitarian paradigm rests on consequentialism and dictates to realize the maximization of the nonmoral value of utility. By contrast, deontology formulates categorical and universal imperatives whose neglect is not permissible in any circumstance. In a highly simplified way, it can be said that the unsolved distortion of utilitarianism is whether the sacrifice of some has to be justified when meant to preserve the lives of many; differently, the unsolved distortion of deontology is the matter of the justifiability of *not* acting in those circumstances in which the forced choice is between a greater or smaller sacrifice of human lives. Somehow utilitarianism leads to controversial dilemmas in the sphere of *acting* whereas deontology leads to similar moral dilemmas in the sphere of *not acting*.

The Rawlsian formulation of neo-contractualism tries to overcome the *impasse* of both theories by establishing, on the one hand, universal and inviolable principles which should provide the moral basis to a fair society and, on the other hand, by indicating the equal distribution of primary goods as the means to realize fairness. Justice, for Rawls, does indeed equal fairness. Central to the theory is therefore the identification of primary goods. This is achieved by means of a rational mechanism named *veil of ignorance* under which moral judges unaware of their racial, health, social, and religious statuses would identify the set of liberties and goods to be universally guaranteed. The rational assumption is that moral judges (whose characteristics are object of a thorough definition; see Rawls 1971) under the condition of not knowing the privileged or disadvantaged position they would have in a real societal setting would identify those liberties and goods that they would regard as essential to the realization of the own self.

When it comes to arguing on the fairness of their distribution, Rawls indicates the rule of “max-min,” following which the *maximum* amount of primary goods which can be *equally distributed* up to the *lowest societal level* acquires the status of the *minimum* amount to be guaranteed in society. It is in this sense that their distribution observes a principle of fairness and that primary goods acquire the status of *rights*. Freedom, for example, is a primary good whose most extensive form (speech, religious beliefs, etc.) should be equally guaranteed up to the most disadvantaged members of society.

The contribution of Moroni to the development of this theory consists of its application to spatial planning theory and, particularly, of the identification of additional primary *spatial* goods. The interesting exercise here is imagining what moral judges would deem as primary and non-violable spatial goods without knowing what their “spatial situation” would be in reality; for example, without knowing their income, their geographical location, and the relevant morphological, climate, and urban conditions. Housing, accessible green areas, access to transportation, and a *safe living environment* are the spatial primary goods that compare in the list compiled by Moroni. It follows that a just spatial planning practice should strive for their equal allocation in society.

Although not explicitly Moroni suggests that *safety* shall be regarded as a primary spatial good. It follows that a minimum level of *spatial safety* corresponds to a right all individuals shall have access to regardless of their spatial condition. This is an interesting conclusion as it provides a point of departure for a new interpretation of safety in the context of spatial planning evaluations. Following the rule of *max-min* the maximum level of spatial safety which can be equally distributed in society up to the lowest societal level acquires the status of the minimum level everybody has to be guaranteed access to. This level should not violate other fundamental rights, for example the right to life and to fundamental freedoms.

It is important to notice that having defined safety as a primary spatial good, and its fair distribution as the aim of the spatial planning practice, implies a “change of scale” in approaching the matter of siting hazardous technologies. The appropriate “where” of risks is no more a problem that regards a given locality, but the locality in relation to the future possible installation of hazardous facilities at regional and national scales. This takes us back to reflect on NIMBY from a totally different perspective. One of the most controversial aspects of the opposition between parties characterized as NIMBY situations is the clash of absolute principles between the right to life of individuals on the one hand and the pursuit of (collective) benefits on the other hand (Peterson and Hansson 2004). Considering that many hazardous establishments (like water treatment plants, energy facilities and energy-related infrastructures, chemical and pharmaceutical factories) produce vital goods to our society, their siting entails the opposition between two equally rightful “goods.” The authors suggest that in order to overcome the clash of absolute principles (and imposing versus opposing the installation of the facility in a legitimate way), so-called residual moral obligations should be observed. These obligations are the obligation to inform, improve, search for knowledge, and compensate, respectively.

It is indubitable that these residual moral obligations should guide morally acceptable siting processes. However when focusing on the single installation as done by Peterson and Hansson (2004) the mentioned unequal distribution of risks, which in the context of site-specific risks can be now rephrased as the matter of the *unequal distribution of spatial safety*, remains unsolved. The added value of a Rawlsian perspective on this matter consists precisely of the addition of the primary moral obligation, for the planning practice, of *distributing spatial safety fairly*. By taking this perspective risk runners acquire the right to oppose a proposed facility when its siting violates such equality, and would instead lose it in all circumstances in which the chosen location respects the principle of a fair spatial safety distribution.

Although debatable, from a spatial planning perspective this is an interesting indication. Coming back to the example of the underground CO<sub>2</sub> disposal in Barendrecht (The Netherlands) it is worth noticing that the research agency assigned to assess the appropriate site among 12 candidate sites focused on the characteristics of each of them according, principally, to their geological characteristics and the technological requirements of the underground disposal. The feasibility study focused on each candidate site as an isolated case, without considering the spatial safety (re)distribution which would have followed the installation of the disposal in each site when put into relation with the broader regional and national scales. More simply, the fact that Barendrecht was the most densely populated area and was already close to major industrial and infrastructural facilities has been overlooked. The fact that (the citizens of) Barendrecht would have been additionally disadvantaged in comparison to other sites has not been regarded as an important decisional element.

Considering the equal distribution of the spatial good of safety is the ideal aim of spatial planning would have instead allowed decision makers to anticipate the situation of (dis) advantage each case would have represented in relation to the wider regional and national context. This outlook does not put the relevance of the safety and technological requirements to be met when choosing the appropriate site for an hazardous installation under discussion; it only suggests to complement these criteria with the one of distributing spatial safety in society as fairly as possible.

Despite these indications are highly theoretical and at the early stage of investigation, they allow to strive for some important conclusions regarding the role of spatial planning processes in relation to technological risks.

## Concluding Remarks: Toward a Moral Understanding of the Relation Between Risks and Space

---

This chapter started with introducing some recent findings in the field of technological risk and spatial planning. The works of Boholm (2004); Boholm and Löfstedt (2004); Owens (2004); and Walker and Simmons (2004) in particular helped in identifying the main areas of investigations to be further explored in order to approach the matter of “siting risks” correctly. In the following, the most controversial aspects of the problem were discussed by referring to the work of other key-scholars in the field, ranging from sociology to ethics and land use scholars. This chapter proposed an ethical outlook on the matter of siting risks by taking together, in particular, ethical and planning theories. Perspectives of further integration will be discussed in the section dedicated to the further research. Hereafter a summary of preliminary conclusions is reported.

The probabilistic and deterministic approaches were identified as the two predominant “school of thoughts” guiding land use planning evaluations in relation to hazardous facilities. The divergences between the two approaches were explained as the result of a non-explicit application of different ethical theories, that are the utilitarian and the deontological theory respectively.

When discussing the matter of framing the “siting risks” issue correctly, that is by overcoming the tendency of interpreting all siting controversies as “not-in-my-backyard” situations, it was proposed that controversies which do evidently lose identifiable spatial and temporal dimensions (that are the *hic et nunc* dimensions) along their development are likely to reveal a more general social concern regarding the technology at issue. It is argued that neglecting the signal value that local oppositions may have in terms of societal rejection of given technological developments is, besides incorrect from an interpretational viewpoint, also highly questionable from a moral one. It is therefore proposed that capturing whether the *hic and nunc* dimensions of siting controversies are definite and identifiable may help in identifying NIMBY (and hence non-NIMBY) cases. Forcing siting processes through the mechanisms of spatial planning regulation and procedures regardless of this distinction does indeed consist of a form of “technological colonization” insensitive to individuals’ values and rights.

The third part of this chapter deepened the reflection on the moral acceptability of hazardous facilities siting by referring, in particular, to the theory of justice of the political philosopher Rawls (1971), the work of the spatial planning theorist Moroni (1997), and the philosophers of risk Peterson and Hansson (2004). This part proposed an interpretation of what was called “spatial safety” as a primary spatial good. By taking this perspective, the aim of the spatial planning practice becomes distributing the maximum possible amount of the primary good of spatial safety in society equally up to the lowest societal level.

This definition of spatial safety has a twofold implication. The first and perhaps most important is that by taking this perspective a morally justifiable imposition of the “here and now” of risks is based on an evaluative approach that puts these two dimensions in relation with the wider spatial and temporal horizons of the social group which has agreed upon the

principle of fairness. These horizons would concretely correspond to the regional or national scales. This means that the “where” of any technological risk should be thought in terms of the situation of spatial safety (re)distribution that the siting would create at regional or national levels. For example, according to the principle of fairness the planner may prevent situations of “risks encroachments” by opting for locating a given hazardous installation in a site wherein the risk profile is lower in comparison to other areas of the region.

The second implication of approaching spatial safety as a primary good regards the moral justifiability of the imposition versus opposition of hazardous facilities in NIMBY situations. The observance of the residual moral obligations of informing, compensating, and improving suggested by Peterson and Hansson (2004) is necessary for empowering actors to impose versus oppose the installation of an hazardous facility. However it is suggested that such observance does not suffice. The right of *opposing* the siting of hazardous installations should be granted to risk runners also in those situations in which they would put into a condition of evident and unnecessary disadvantage in comparison to other areas of their region or nation. In other words imposing “risks on risks” in certain areas due, for example, to the presence of existing infrastructures that facilitate the further installation of hazardous facilities should be regarded as unfair from a distributive justice perspective. Rather than being concentrated spatial *un-safety* should be distributed; hence, opposing the installation of hazardous facilities on the ground of unfairness should be regarded as legitimate.

These conclusions are highly general and still in their embryonic formulation. However they sign a promising research trajectory, whose outline concludes the present work.

## Further Research

---

Each part of this chapter presented the problem of “siting risk” under a policy, spatial and risk analysis light to end up with discussing its most problematic ethical implications. This is not casual. In front of technological risks, and more generally in front of the irreducible complexity of facing uncertainty, ultimately the discussion gravitates around precise, although often not explicit, ethical considerations.

This premise serves to point out that the further research needed to tackle the matter of siting hazardous technologies in a sustainable and technologically, socially, and morally acceptable way passes through the explicit adoption of precise ethical principles. As discussed along this chapter, different ethical theories may lead to substantially different approaches to the governance of risks in society and, in the specific case of site-specific technologies, to assessing their “where.” It is hence toward ethics that spatial planning should orient its vision when designing evaluative processes sensitive to the aim of achieving a morally acceptable relation between risks and space. This implies opening the spatial planning research to the inputs provided by the ethics of technological risks literature and striving for a solid synergy between the practice of spatial design and the application of (past and current) ethical theories.

This indication opens to two distinct research horizons. The one discussed in this work relates to the matter of siting risks by adopting a distributive justice approach. One issue that remains to be investigated in this regard is which is the level of the primary spatial good of safety that should be minimally guaranteed in society. As mentioned above this level should correspond to the one that does not violate the primary right to life. But this is evidently a deadlock, as there is no risk which per se does not threaten it.

Taking a neo-contractualist perspective could again help in solving the *impasse*. The challenge is imaging what moral judges would regard as the non-violable level of spatial safety (and hence as the acceptable level of risk) each member of society should be guaranteed regardless of her spatial condition. Intuitively in the era of technological dependency and constant exposure to involuntary (yet somehow necessary) risks this is not a straightforward answer. It is suggested that put in front of the dichotomy between determinism and probabilism moral judges would hang in favor of a cautious consideration of the latter. When their objects are hazardous technologies of proven necessity to sustain one's health and capabilities, rigorous deterministic considerations of the relevant risks are indeed poorly justifiable. It is suggested that it exists as a fundamental moral intuition that dictates to allocate societal resources toward the prevention of the risks associated to vitally beneficial technologies without cultivating the ambition of taking them to zero. The caution consists precisely in remembering that "quantity is the trap of all traps" when it comes to define an acceptable versus non-acceptable risk level. Paradoxically, without incorporating individuals' intangible values within the very characterization of risks the matter remains at the abstract level of calculating probabilities  $\times$  consequences. The question of what is the minimum level of spatial safety to be fairly distributed in society is legitimate only if safety is not regarded as the result of a mono-criterion assessment, but also as what individuals' regard as the nonnegotiable space values which are at the basis of their well-being.

The second research horizon indicated by this discussion developed along this chapter is of more general relevance and regards the need of making the ethical theories informing spatial planning theories explicit together with the moral implications of concrete spatial planning decisions. It is suggested that in a world of increasing complexity, but also of renewed attention for the most fundamental human values, there is no spatial planning decision that does not require precise ethical considerations. Planning the space equals designing our "present future" and determining our condition of liberty and equality into it; in this respect, design can be considered as the *longa manu* of ethics. This implies a fundamental responsibility for the spatial planning discipline at all levels; and that is a moral responsibility.

## References

- Adams M (2002) The precautionary principle and the rhetoric behind it. *J Risk Res* 5:301–316
- Ale BMJ (2005a) Tolerable or acceptable: a comparison of risk regulation in the United Kingdom and in the Netherlands. *Risk Anal* 25(2):231–241
- Ale BMJ (2005b) Living with risk: a management question. *Reliab Eng Syst Saf* 90:196–205
- Amendola A (2001) Integrated risk management: recent paradigms and selected topics, Integrated Management for Disaster Risk, research booklet no. 2. Disaster Prevention Research Institute, Kyoto University, Japan
- Amendola A, Contini S, Ziomas I (1992) Uncertainty in chemical risk assessment: results of a European benchmark exercise. *J Hazard Mater* 29: 347–363
- Apostolakis G (1990) The concept of probability in safety assessment of technological systems. *Science* 250: 1359–1364
- Apostolakis G (2004) How useful is quantitative risk assessment? *Risk Anal* 24:515–520
- Arcuri A (2005) Governing the risk of ultra-hazardous activities: challenge for contemporary legal systems. Ph.D. thesis, Rotterdam Erasmus University
- Asveld L, Roeser S (2009) The ethics of technological risk. Earthscan, London
- Basta C (2009) Risk, territory and society: challenge for a joint European regulation. Ph.D. thesis, Delft University of Technology
- Basta C, Neuvel J, Zlatanova S (2006) Risk-maps informing land use planning processes. *J Hazard Mater* 145(1–2):241–249

- Basta C, Struckl M, Christou MD (2008) Implementing art. 12 of the Seveso II directive: overview of roadmaps for land use planning in selected member states. JRC technical report, EUR23519 EN
- Beck U (1986:1992) Risk society: towards a new modernity. Sage, London
- Boholm A (2004) What are the new perspectives on siting controversies? *J Risk Res* 7(2):99–100
- Boholm A, Löfstedt R (2004) Facility siting: risk, power and identity in land use planning. Earthscan, London
- Bottelberghs PH (2000) Risk analysis and safety policy developments in The Netherlands. *J Hazard Mater* 71:117–123
- Breunese JN, Remmelt G (2009) Inventory of potential locations for demonstration project CO<sub>2</sub>-storage, TNO-034-UT-2009-02024, The Netherlands
- Brown G, Raymond C (2007) The relationship between place attachment and landscape values: toward mapping place attachment. *Appl Geogr* 27(2):89–111
- Christou MD (1998) Consequence analysis and modeling. In: Kirchsteiger C (ed) Risk assessment and management in the context of the Seveso II directive. Elsevier, Amsterdam
- Christou MD, Amendola A (1998) How lessons learned from benchmark exercises can improve the quality of risks studies. In: Mosleh A, Bari RA (eds) Proceedings of probabilistic safety assessment and management PSAM 4, vol 2. Springer, New York, pp 840–845
- Christou MD, Mattarelli M (2000) Land-use planning in the vicinity of chemical sites: risk-informed decision-making at a local community level. *J Hazard Mater* 78:191–222
- Christou MD, Struckl M, Biermann T (2006) Land use planning guidance in the context of article 12 of the Seveso II directive 96/82/EC, etc., online. [http://ec.europa.eu/environment/seveso/pdf/landuseplanning\\_guidance\\_en.pdf](http://ec.europa.eu/environment/seveso/pdf/landuseplanning_guidance_en.pdf). Accessed June 2011
- Council Directive 2003/105/EC of the European Parliament and of the Council of 16 Dec 2003, OJ L 345/97
- Council Directive 82/501/EEC on the major-accident hazards of certain industrial activities, OJ L 230, 5 Aug 1982
- Council Directive 96/82/EC on the control of major-accident hazards involving dangerous substances, OJ L 10/13
- Cozzani V et al (2001) The use of quantitative area risk assessment techniques in land-use planning. In: Papadakis GA (ed) Risk management in the European Union of 2000, EUR 19664, EN. Commission of the European Communities, Brussels
- Cozzani V et al (2006) Application of land-use planning criteria for the control of major accident hazards: a case-study. *J Hazard Mater* 136:170–180
- Daamen DDL et al (2010) Wat weten en vinden Barendrechters van het CO<sub>2</sub> opslag plan en van voorlichting en besluitvorming over dit plan? Resultaten van een enquête in mei 2010 onder ruim 800 inwoners, Leiden University, online. <http://www.co2-cato.nl/cato-2/publications/publications/wat-weten-en-vinden-barendrechters-van-het-co2-opslag-plan-en-van-voorlichting-en-besluitvorming-over-dit-plan>
- De Marchi B, Funtowicz S, Ravetz J (1996) Seveso: a paradoxical classical disaster. In: Mitchell J (ed) The long road to recovery: community responses to industrial disaster. United Nation University Press, Tokyo
- Ersdal G, Aven T (2008) Risk informed decision-making and its ethical basis. *Reliab Eng Syst Saf* 93: 197–205
- Golany GS (1995) Ethics and urban design: culture, form and environment. Wiley, New York
- Hague C, Jenkins P (2005) Place identity, participation and planning. Routledge, London
- Hansson SO, Peterson M (2001) Rights, risks, and residual obligations. *Risk Decis Policy* 6:157–166
- Hare MR (1999) What are cities for? The ethics of urban planning. In: Hare MR (ed) Objective prescriptions and other essays. Oxford University Press, Oxford
- Hewitt K (1997) Regions of risk. Addison Wesley Longman, Essex
- Hillerbrand R (2010) On non-propositional aspects in modelling complex systems. *Analyse und Kritik* 32:107–120
- Horlick-Jones T (1998) Social theory and the politics of risk. *J Contingencies Crisis Manag* 6:64–67
- Jongejan RB (2008) How safe is safe enough?. Ph.D. thesis, Delft University of Technology
- Judge Asquit (1949) Edwards v. the National Coal Board, All England Law Reports, vol 1
- Keith M, Pile S (1993) Place and the politics of identity. Routledge, London
- Lauridsen K et al (2002) Assessment of uncertainties in risk analysis of chemical establishments. The ASSURANCE project. Final summary report. Risø-R-1344(EN)
- Lees FP (1996) Loss prevention in the process industry. Butterworth-Heinemann, Oxford
- Lesbirel S-H, Shaw D (2005) Managing conflicts in facility siting: An international comparison. Edward Elgar Publishing Limited, Cheltenham, UK
- Möller N, Hansson SO, Peterson M (2006) Safety is more than the antonym of risk. *J Appl Philos* 23(4): 419–423
- Moroni S (1993) Planning theory, practical philosophy and phronesis. *Comments on Flyvbjerg. Plann Theory* 9:120–133

- Moroni S (1994) *Territorio e Giustizia Distributiva* (English trans: Territory and distributive justice). Franco Angeli, Milan
- Moroni S (1997) *Etica e Territorio* (English trans: Ethics and land use). Franco Angeli, Milan
- Moroni S (2004) Towards a reconstruction of the public interest criterion. *Plann Theory* 3(2):151–171
- Moroni S (2006) Planning, evaluation and the public interest. In: Alexander ER (ed) *Evaluation in planning: evolution and prospects*. Ashgate, Aldershot, pp 55–71
- Owens S (2004) Siting, sustainable development and social priorities. *J Risk Res* 7(2):101–114
- Pellizzoni L, Ungaro D (2000) Technological risk, participation and deliberation. Some results from three Italian case studies. *J Hazard Mater* 78: 261–280
- Peterson M (2001) New technologies and the ethics of extreme risks. *Ends Means* 5:22–30
- Peterson M (2003) Risk, equality, and the priority view. *Risk Decis Policy* 8:17–23
- Peterson M, Hansson SO (2004) On the application of right-based moral theories to siting controversies. *J Risk Res* 7(2):269–275
- Pettit P (1991) Consequentialism. In: Singer P (ed) *A companion to ethics*. Blackwell Publishers, Oxford
- Rawls J (1971) *A theory of justice*. Harvard University Press, Cambridge, MA
- Renn O (1992) Concepts of risk: a classification. In: Krinsky S, Godling D (eds) *Social theories of risk*. Praeger Eds, Westport
- Renn O (2006) Participatory processes for designing environmental policies. *Land Use Policy* 23:34–43
- Renn O, Graham B (2005) White paper on risk governance. International Council on Risk Governance (eds), Geneva
- Roeser S (2006) The role of emotions in judging the moral acceptability of risks. *Saf Sci* 44:689–700
- Roeser S (2010) Emotions and risky technologies. Springer, Berlin
- Salvi O et al (2005) Toward an integrated approach of the industrial risk management process in France. *J Loss Prev Process Ind* 18:414–422
- Schmidt-Thomé P (2006a) Integration of natural hazards, risk and climate change into spatial planning practices. Ph.D. thesis no.193 of the Department of Geology, University of Helsinki
- Schmidt-Thomé P, Kallio H (2006b) Natural and technological hazard maps of Europe. In: Schmidt-Thomé P (ed) Geological survey of Finland, Special Paper no 42
- Schmidt-Thomé P et al (2006c) The spatial effects and management of natural and technological hazards in Europe, deliverable 1.3.1, ESPON project
- Schrader-Frechette KS (1991) Risk and rationality: philosophical foundations for populist reform. University of Berkeley, California
- Schütz H, Wiedemann PM (1995) Implementation of the Seveso directive in Germany: an evaluation of hazardous incident information. *Safety Sci* 18(3):203–214
- Shrivastava P (1992) *Bhopal: anatomy of a crisis*. Paul Chapman, London
- Shrivastava P (1996) Long-term recovery from the Bhopal crisis. In: Mitchell JK (ed) *The long road to recovery: community responses to industrial disasters*. United Nation University Press, Tokyo
- Simmons P, Walker G (2004) Living with technological risk: industrial encroachment on sense of place. In: Boholm A, Loftsdóttir R (eds) *Facility siting: risk, power and identity in land use planning*. Earthscan, London
- Slovic P (1991) Beyond numbers: a broader perspective in risk communication and risk perception. In: Mayo DG, Hollander D (eds) *Acceptable evidence: science and values in risk management*. Oxford, New York
- Slovic P (1999) Trust, emotions, sex, politics and science: surveying the risk-assessment battlefield. *Risk Anal* 19:689–701
- Slovic P (2002) Perception of risk posed by extreme events. Paper presented at the congress risk management strategies in an uncertain world, Palisades, New York, 12–13 Apr
- Sustein CR (2005) *Laws of fear: beyond the precautionary principle*. Cambridge University Press, Cambridge
- Taebi B (2010) Sustainable energy and the controversial case of nuclear power. In: Raffaele R, Robson W, Selinger E (eds) *Sustainability ethics: 5 questions*. Automatic Press, Copenhagen
- Taebi B, Kadak AC (2010) Intergenerational considerations affecting the future of nuclear power: equity as a framework for assessing fuel cycles. *Risk Anal* 30(9):1341–1362
- Taibakhsh K (2001) *The promise of the city: space, identity and politics in contemporary social thought*. University of California Press, California
- Vermaas PE et al (2008) *Philosophy and design*. Springer, Dordrecht
- Walker G (1991) Land use planning and industrial hazards. A role for the European community. *Land Use Policy* 8(3):227–240
- Walker G (1995) Land use planning, industrial hazards and the COMAH directive. *Land Use Policy* 12(3):187–191
- Walker G et al (1999) Risk communication, public participation and the Seveso II directive. *J Hazard Mater* 65:179–190
- Weick K (1987) Organizational culture as source of high reliability. *Calif Manage Rev* 29:112

# 12 Intergenerational Risks of Nuclear Energy

Behnam Taebi

Delft University of Technology, Delft, The Netherlands

<i>Introduction</i> .....	296
<i>Nuclear Fuel Cycles and the Issue of Waste</i> .....	297
<i>History of Intergenerational Equity in Nuclear Waste Management</i> .....	299
Waste Management Policies: “a Desire for Equity” .....	301
Safety for People of the Future .....	302
Security for People of the Future .....	302
Equal Opportunity: Retrievable Disposal .....	304
<i>Policy-Making and the Principle of Diminishing Responsibility</i> .....	304
Geological Disposal and Long-Term Uncertainties .....	306
<i>State of the Art in Technology: Challenge to Geological Disposal</i> .....	308
<i>Four Counterarguments to the Feasibility and Desirability of P&amp;T</i> .....	311
<i>Conclusions</i> .....	313
<i>Further Research</i> .....	314
<i>Notes</i> .....	315

**Abstract:** Nuclear energy is one of the clearest examples of a technology that brings about risks beyond generational borders. These risks emanate particularly from nuclear waste that needs to be isolated from the biosphere for very long periods of time. Principles of intergenerational equity currently underlie waste management policies, arguing that we should not impose *undue burdens* on future generations. This chapter scrutinizes the way in which such intergenerational equity principles deal with the issue of long-term risks.

The present consensus within the nuclear community is that nuclear waste should be buried in geological repositories rather than kept in surface storage places. This is particularly based on the notion that repositories are believed to be safer in the long run. Such long-term safety seems to be disputable as it relies on great long-term uncertainties which, in turn, necessitate sanctioning a distinction between different future people. Putting distant future generations at a disadvantage does, however, lack solid moral justification, which should urge us to reconsider our temporal moral obligations in the light of recent technological developments. The technological possibility of substantially reducing the waste lifetime through Partitioning and Transmutation (P&T) is believed to challenge geological disposal, thus placing long-term surface storage in a new perspective. P&T is, however, a laboratory-scale technology which means that substantial investment will be required before industrial deployment can take place. Moreover, the deployment of this technology creates additional safety risks and economic burdens for the present generation. Nevertheless, the potential possibility to diminish “*undue burdens*” for future generations is too relevant to be neglected in discussions on nuclear waste management policies. The question that will furthermore be explored is to what extent should we rely on future technological possibilities in today’s policy-making?

## Introduction

---

The rapidly growing energy consumption level, future forecasts, and climate change have prompted a new debate on alternative energy resources. Alongside green energy such as wind and solar power, a new nuclear era seems to have dawned. According to the World Nuclear Association, there were 443 operational nuclear reactors in March 2011. In addition, a further 62 reactors are currently under construction, 158 have been ordered or planned, and 324 have been proposed ([WNA 2011](#)). Nuclear energy now accounts for almost 16% of all the electricity produced worldwide.

The main advantage of nuclear energy – when compared to fossil fuels – is that it can produce a large amount of energy from relatively small amounts of fuel while generating very low greenhouse gas levels. However, there are serious drawbacks attached to nuclear energy production such as those surrounding safety concerns in reactors; the recent series of accidents in the Fukushima Nuclear power plants in Japan serve to heighten such concerns. In addition, there are the proliferation threats if the technology is abused by being employed for destructive purposes and, as always, there is the issue of how to deal with nuclear waste. Nuclear waste remains radiotoxic for a long period of time before decaying to a nonhazardous level, a level defined by the radiotoxicity of the same amount of uranium ore. This period is known as the waste lifetime and for the remaining materials in a once-through fuel cycle that amounts to 200,000 years. Recycling technologies (reprocessing) are capable of reducing this waste lifetime to 10,000 years. Recent developments in nuclear waste management – Partitioning and

Transmutation (P&T) – demonstrate at laboratory level that it is possible to reduce the waste lifetime yet further, to a couple of hundred years (NRC 1996).

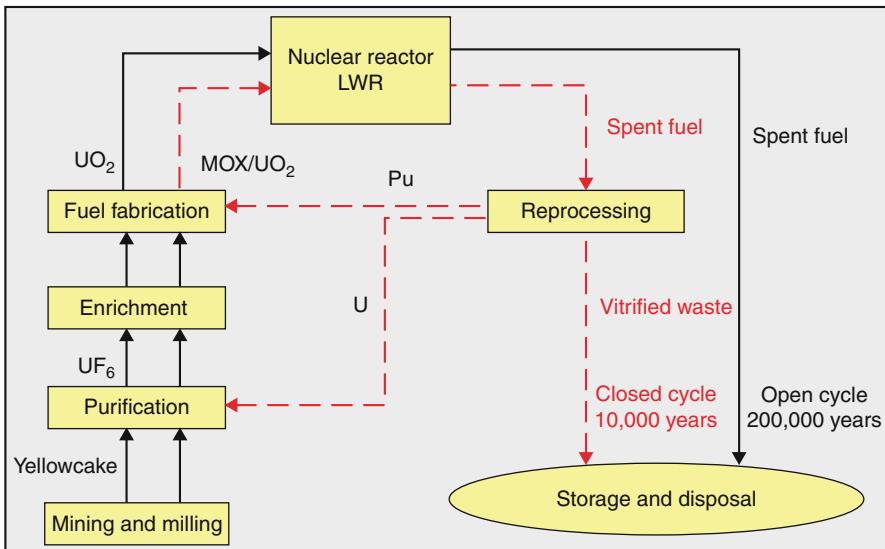
For the current and future protection of human health and the environment when it comes to dealing with radioactive waste, the International Atomic and Energy Agency (IAEA) has laid down certain principles for Radioactive Waste Management, one of which states that nuclear waste should be managed in such a way that it “will not impose undue burdens on future generations” (IAEA 1995, Pr. 5). This “undue burdens” clause could best be situated within the framework of intergenerational equity or equity across generations (NEA-OECD 1995). Current policy in nuclear energy-producing countries stems mainly from intergenerational equity considerations, and from the striving for an equitable distribution of risks and burdens inspired by the belief that the safety and security of future generations should not be jeopardized. Intergenerational equity also involves guaranteeing equal opportunities for future generations. In nuclear energy-related discussions, equal opportunity mainly pertains to the *retrievability* of waste and to its potential economic value.

The consensus within the nuclear community for the ultimate disposal of waste seems to be in favor of burying waste in geological repositories rather than storing it on the surface; this consensus is founded on the long-term safety (and security) assurances supposedly guaranteed by host geological formations (IAEA 2003). However, long-term safety depends on certain considerable uncertainties, which necessitate sanctioning a distinction between different future generations. I argue that this distinction lacks moral justification and that it would therefore be best for us to avoid these uncertainties. Implementing P&T allows for the latter, as the period of necessary care for P&T waste is substantially shorter.

This chapter is organized as follows. In the next section, I briefly discuss the production of nuclear energy and the different options for the final isolation of waste. In the section **● History of Intergenerational Equity in Nuclear Waste Management**, I elaborate on the way in which the notion of intergenerational equity has influenced nuclear waste management policies. The section **● Policy-making and the Principle of Diminishing Responsibility** scrutinizes the notion of “diminishing responsibility over the course of time” as a leading notion in nuclear waste policies. I then examine the moral legitimacy of distinguishing between future people, particularly when designing geological repositories. The state of the art in technology is discussed in the section **● State of the Art in Technology: Challenge to Geological Disposal**. The section **● Four Counterarguments to the Feasibility and Desirability of P&T** puts forward four possible counterarguments to the application of P&T. The concluding two sections present the findings in brief and highlight the further research possibilities.

## Nuclear Fuel Cycles and the Issue of Waste

Before moving on to the discussions on how policy-making has been affected by considerations of justice to future generations, let me first quickly present the methods for the production of nuclear power and elaborate on the question of how different waste types are dealt with. Nuclear energy is produced in a *nuclear fuel cycle* that starts with the mining of uranium (U), the fabrication of fuel and irradiation in a nuclear reactor (the front-end stage), and finishes with the possible treatment of irradiated fuel emanating from the reactor before ending with the final disposal of remaining waste (the back-end stage); see **● Fig. 12.1**.

**Fig. 12.1**

The open and closed nuclear fuel cycles, from uranium ore to final disposal

Irradiated fuel is referred to as *spent fuel* rather than waste because the way in which spent fuel is dealt with represents a crucial choice in terms of nuclear waste management. The first production method, or fuel cycle, is one in which irradiated fuel is viewed as waste and therefore has to be isolated from the biosphere for a long period of time, thus creating an *open fuel cycle*. The irradiating of uranium (U) produces other materials, including plutonium (Pu), which is a very long-lived radioactive isotope. Apart from plutonium, other residual radioactive materials, *minor actinides*, as well as *fission products* will be formed. Actinides are elements with similar chemical properties. Uranium and plutonium are the major constituents of spent fuel and so they are known as major actinides. Neptunium, americium, and curium are produced in much smaller quantities and are thus termed minor actinides. Fission products are a mixture of radionuclides that will decay to a nonhazardous level after approximately 250 years. The presence of major actinides in spent fuel defines the waste lifetime in an open fuel cycle; neither minor actinides nor fission products have a significant effect on long-term radiotoxicity. The waste lifetime of spent fuel in such a fuel cycle is 200,000 years and it is dominated by plutonium.

The second method is one in which we “destroy” or convert the very long-lived radionuclides to shorter-lived material in accordance with a *closed fuel cycle*. Removing plutonium from spent fuel can substantially diminish the waste lifetime. In a closed fuel cycle, plutonium and any remaining uranium are isolated and recovered during a chemical treatment phase which is referred to as *reprocessing*. Reprocessed uranium could be added to the beginning of the fuel cycle and both uranium and plutonium could be used to produce *mixed oxide fuel*, a combination of uranium oxide and plutonium oxide that can be used in nuclear reactors as a fuel (Wilson 1999). The closed fuel cycle waste stream is referred to as high-level waste (HLW) or *vitrified waste* and it has a waste lifetime of approximately 10,000 years.

Irrespective of the fuel cycle choice, the waste remaining after optional treatment needs to be disposed of. In waste management, a distinction is made between storage and disposal: storage entails keeping the waste in purpose-built facilities above ground or at a certain depth beneath the surface, while disposal entails isolating and depositing waste at significant depths (of several hundred meters) below the surface in engineered facilities. The latter are termed geological repositories or simply repositories.

Spent fuel is usually stored under water for a period of time – varying from a couple of years to several decades – after having been removed from the reactor core; this stage is called the *interim storage* stage. Water serves as a radiation shielding and cooling fluid; heat exchange facilities will remove heat from the circulating cooling water (IAEA 1998). The interim storage of waste is also a crucial factor in the safe management of radiotoxic waste, since it is designed to allow radioactive decay to reduce the level of radiation and heat generation before the final disposal stage. It is especially heat generation that very much influences the capacity of a repository (Bunn et al. 2001).

A commonly proposed alternative to geological disposal is long-term monitored surface storage. However, the technical community largely appears to disregard this option since it sees surface storage as merely an interim measure prior to disposing of waste in geological repositories (IAEA 2003; NEA-OECD 1999). Up until now, all the available facilities for spent fuel and high-level waste tend to have been located above ground or at very shallow depths, designed for interim storage. Some people are, however, concerned that this interim storage phase may well become, *de facto*, perpetual.

Depositing the waste in space or disposing of it in inaccessible deep-sea sediments – like for instance beneath the Antarctic ice – are the alternatives that have been proposed (KASAM 2005, Ch. 3). These options are not, however, being taken seriously, due to the unacceptable safety risks involved together with what would amount to the violation of international conventions.

## History of Intergenerational Equity in Nuclear Waste Management

Widespread concerns about depleting the Earth's resources and damaging the environment have recently triggered new debate on the equitable sharing of goods over the course of generations or, on *intergenerational justice*.<sup>1</sup> This concept of justice was first introduced by John Rawls (1971) who alluded to intergenerational distributive justice.

Intergenerational justice in connection with climate change has received increasing attention in the literature in recent years (Page 1999; Gardiner 2001; Athanasiou and Baer 2002; Shue 2003; Gardiner 2003, 2004; Meyer and Roser 2006; Page 2006; Gardiner 2006a; Caney 2006). In nuclear waste management this notion of justice or equity<sup>2</sup> across generations has been influential, particularly in promoting geological repositories as final disposal places for nuclear waste; later in this section I will elaborate on this issue. Before that I would like to pause for a moment to reflect on the claim that there is a problem of intergenerational justice that emanates from nuclear power production.

I shall adhere here to Stephen Gardiner's discussions about "The Pure Intergenerational Problem" (PIP), in which he imagines a world consisting of temporally distinct groups that can asymmetrically influence each other; "earlier groups have nothing to gain from the activities or attitudes of later groups." Each generation has access to a diversity of temporally diffuse commodities. Engaging in activity with such goods culminates in modest present benefits

and substantial future cost and that, in turn, poses the problem of fairness. Gardiner refers to the problem of energy consumption and anthropogenic carbon dioxide which causes climate change and has predominantly good immediate effects but deferred bad effects. Even if present benefits exceed future costs (assuming that we can compare these entities), this is a moral problem because just as with most theories of justice, distribution is independent of overall utility (Gardiner 2003, pp. 483–488).

There are two ways in which PIP relates to nuclear power production and to waste management. First of all, assuming that this generation and those that immediately follow will continue depleting uranium, a nonrenewable resource, there will be evident intergenerational equity considerations to bear in mind. Secondly, the production of nuclear waste, and its longevity in terms of radioactivity, signifies substantial present benefits with deferred costs, as stated in Gardiner's PIP.

Another cause for concern in nuclear energy-related discussions is our beneficial temporal position with regard to successive generations: “our temporal position allows us to visit costs on future people that they ought not to bear, and to deprive them of benefits they ought to have”; Gardiner refers to this as “The Problem of Intergenerational Buck Passing” (Gardiner 2006b, p. 1). As we can reasonably expect that the incentive structure remains the same for all generations, the Problem of Intergenerational Buck Passing will be exacerbated over the course of time, all of which gives rise to moral justification for limiting the impacts of our actions that have intergenerational consequences (Gardiner 2006b, pp. 2–3).

I argue that the intergenerational problem resulting from nuclear power production entails certain moral obligations and that contemporaries must not endanger the interests of future generations. These obligations can manifest themselves in two different ways. Firstly, there is the depletion of a resource, namely, uranium, that will not renew itself for future generations which justifies placing certain restrictions on depleting these resources.<sup>3</sup> Assuming that well-being significantly relies on the availability of energy resources – a claim that could be historically underpinned by considering developments from the time of the industrial revolution up until the present – we could be said to have an obligation to ensure well-being for the future. Another, perhaps more important obligation, relates to the longevity of nuclear waste and to the fact that its inappropriate burial can harm future generations. So the next moral obligation that the intergenerational problem creates is that of not harming future people.

There are, however, certain theoretical and practical objections to this reasoning. The theoretical problem is whether rights can be ascribed to future people to justify obligations and whether we can harm future people whose very existence depends on our actions and inaction. At the practical level of applying these ideas, we might consider the question of how far into the future these obligations extend and how we can deal with the uncertainties linked to such long-term predictions. In this chapter, I take the liberty of not extensively discussing such theoretical impediments. In the following paragraphs, I shall touch on the theoretical objections and briefly argue why I do not find them persuasive enough to free us from talking about the obligation not to harm future generations. In the next section, more attention will be given to the relevance and legitimacy of a distinction between different future generations in the light of long-term uncertainties.

Future people's identity and numbers are very much contingent on the actions and policy choices we make now. As it is the moment of conception that determines which individuals will come into existence and as different policy choices result in different individuals, we can never be accused of harming future individuals since, in line with Parfit's (1983) non-identity

problem, it is changing policy that will change the number and identity of all these still to be born individuals. The non-identity problem is a serious theoretical problem that theories of intergenerational justice, particularly the personal and harm-based theories, are facing. Certain solutions have been proposed, but they are either not definite or have a too small scope; the relevant philosophical inquiries should certainly continue to fix this theoretical ineptitude. The non-identity problem, however, does not reduce our moral obligations to posterity. This is a counterintuitive conclusion that is gladly supported by just a few people, Schwartz (1978) being one of them. It is also worth noting that Parfit (1984, p. 357) would disagree with such a conclusion as he states that the mere fact that we can influence the well-being of future generations creates a certain obligation toward posterity.

The next problem is whether we have any obligations toward these contingent people and whether these obligations are founded in rights. Some scholars argue that “the ascription of rights is properly to be made to actual persons – not possible persons” (Macklin 1981, p. 151); see also Beckerman and Pasek (2001). Here, I follow the interest theory of rights argumentation to the effect that if agent X has a right then that implies that “other things being equal, like aspects of X’s well-being (his interest) is sufficient reason for holding some other person(s) to be under a duty” (Raz 1986, p. 183). We can safely assume that there will be future generations and that these people will have interests which will depend on the actions of the current generation. “The identity of the owners of these interests is now necessarily obscure, but the fact of their interest-ownership is crystal clear” (Feinberg 1981, p. 148).

To conclude, the depletion of uranium and the longevity of nuclear waste cause the problem of intergenerational justice and that, in turn, creates moral obligations for present generations to ensure future people’s well-being – in terms of resource availability – and not to harm future people by inadequately burying nuclear waste. It should be clear that I am not intending to examine the extent of the stringency of these moral obligations nor to elaborate on any theoretical impediments that may arise (elsewhere I take up this challenge (Taebi 2011)). In this chapter, I am merely assuming that the production of nuclear power creates certain moral obligations, the extent of which remains a subject of ongoing discussion.

### Waste Management Policies: “a Desire for Equity”

Having discussed the moral obligations that ensue from the intergenerational problem that the production of nuclear energy creates, I shall now return to the nuclear waste management principles and to the overarching notion of intergenerational equity. The long-term concerns, as outlined above, have triggered a debate on how to deal with radiotoxic waste in an equitable way. The level of acceptance<sup>4</sup> of risks for present generations is proposed as a reasonable indication for the future. The International Atomic and Energy Agency (IAEA 1995) laid down several principles of radioactive waste management, in which concerns about the future were expressed in terms of the “achievement of intergenerational equity.” It was asserted that nuclear waste should be managed in such a way that it “will not impose undue burdens on future generations” (IAEA 1995, Pr. 5). The Nuclear Energy Agency (NEA) reiterated those principles in a Collective Opinion, which stated that geological disposal should be preferred to aboveground storage on the basis of considerations of intergenerational equity: “our responsibilities to future generations are better discharged by a strategy of final disposal [underground] than by reliance on [above ground] stores which require surveillance, bequeath

long-term responsibility of care, and may in due course be neglected by future societies whose structural stability should not be presumed” (NEA-OECD 1995, p. 5). All national programs have already subscribed to the concept of geological disposal as a “necessary and a feasible technology”; but some countries prefer to postpone implementation in order to first evaluate other options and alternatives (NEA-OECD 1999, p. 11).

In the following paragraphs, I will present current thinking on waste management policies in terms of the underlying philosophical and ethical considerations stemming from the principle of intergenerational equity. The basic notion is that the present generation is required to ensure that there is an equitable distribution of risks and burdens, which must ensure the safety and security of future people. In addition, equity across generations also involves the assurance of equal opportunities for future generations when dealing with nuclear waste. Three ethical values relevant to current nuclear waste policy – namely those of safety, security, and equal opportunity – will be reviewed below. I will furthermore focus on the issue of how these principles motivate geological disposal, as opposed to aboveground storage.

## Safety for People of the Future

From the early days of nuclear energy deployment, the safety of future generations has been a primary concern, as can be concluded from the guidelines laid down in 1955 by the US National Academy Committee on the Geological Aspects of Radioactive Waste Disposal (NRC 1966). In spite of this early recognition, it was a long time before the nuclear community explicitly mentioned the safety of future people as a concrete concern. In 1984, the Nuclear Energy Agency first pronounced “a desire for equity” and acknowledged a need for “the same degree of protection” for people living now and in the future (NEA-OECD 1984). The IAEA articulates these concerns in its Safety Principles where it states that nuclear waste should be managed in such a way that “predicted impacts on the health of future generations will not be greater than relevant levels of impact that are acceptable today” (NEA-OECD 1995, p. 6) and it refers to this as the neutrality criterion. Geological disposal is believed to ensure safety as it is seen as a resistance to containment of the waste over very long periods of time. The engineered facility, together with the natural safety barrier of the host geological formation must guarantee that “no significant radioactivity” will even return to the surface environment (NEA-OECD 1999, p. 11).

## Security for People of the Future

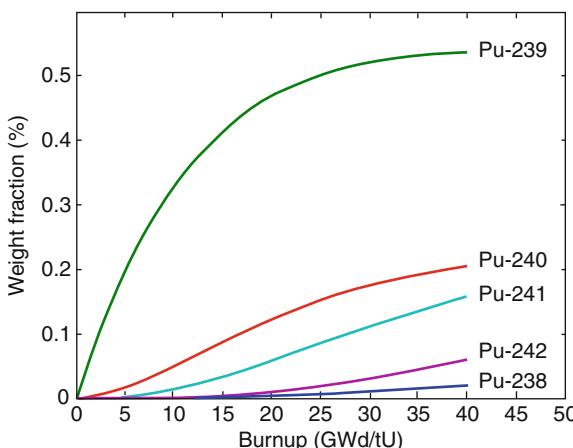
“[T]he same degree of protection” as that stated by the NEA (1984) not only refers to public health issues, but also to future security concerns. Security relates to the unauthorized possession or theft of radiotoxic waste in order to either sabotage or use these materials for the production of nuclear weapons (IAEA 2007, pp. 133–134). The main concerns relate to the threat of nuclear weapon proliferation which is extremely relevant given the current state of affairs in the world. Proliferation threats arise either from the using of highly enriched uranium (HEU) which has been enriched up to 70% (and higher) or from the production or separation of plutonium. Hundreds of tons of highly enriched uranium and weapon-grade plutonium derived from the dismantled nuclear weapons found in American and Russian stockpiles are

the “deadly legacy” of the Cold War that give rise to so much concern (Bunn 2000). Apart from deriving from disarmed nuclear warheads, both highly enriched uranium and plutonium can also be produced using the technology currently available in many nuclear energy-producing countries. As soon as uranium becomes more than 20% enriched, the intentions are evidently for destructive ends; such action in any declared facility is immediately detected by the IAEA. Enrichment facilities are not present in all nuclear energy-producing countries.

Plutonium, on the other hand, is produced during fuel irradiation and separated during reprocessing in countries favoring the closed fuel cycle approach. The extracted plutonium is destined for use as a fuel ingredient (as mixed oxide fuel), but it also carries proliferation threats. To illustrate the seriousness of these potential risks: 8 kg of *weapon-grade* plutonium ( $^{239}\text{Pu}$ ) is sufficient to produce a bomb with the devastation potential of the Nagasaki bomb. The kind of plutonium which, under normal circumstances emerges from a power reactor consists of different isotopes including  $^{238}\text{Pu}$ ,  $^{240}\text{Pu}$ , and  $^{239}\text{Pu}$ ;  $\circlearrowleft$  Figure 12.2 shows the buildup of different plutonium isotopes during energy production or fuel irradiation. When more than 93% of  $^{239}\text{Pu}$  is present, plutonium becomes a weapon-grade element and below 80% of this isotope it is referred to as *reactor-grade* or *civilian* plutonium. For destructive purposes, plutonium must contain as much as possible  $^{239}\text{Pu}$  in proportion to the relatively short burnup time, as can be seen in  $\circlearrowleft$  Fig. 12.2.

To conclude, “[d]eploying reactor-grade Pu is less effective and convenient than weapon-grade in nuclear weapons”, but still “[...] it would be quite possible for a potential proliferator to make a nuclear explosive from reactor-grade plutonium using a simple design that would be assured of having a yield in the range of one to a few kilotons and more, using an advanced design” (DOE 1997). Separated plutonium whether it is weapon-grade or reactor-grade carries serious security risks.

Let us now move on to the question of how these considerations relate to the choice of final disposal waste methods. Geological repositories are believed to ensure security which is perceived as “resistance to malicious or accidental disturbance [...] over very long times”



$\blacksquare$  Fig. 12.2  
Burnup of different plutonium isotopes in a Light-Water reactor<sup>5</sup>

better than easily accessible above ground storage facilities (NEA-OECD 1999, p. 11). “[W]aste stores [on the surface] are vulnerable to inadvertent or deliberate intrusion by humans if not kept under close surveillance. This places obligations on future generations” (IAEA 2003, p. 5). The IAEA (2003, p. 7) further asserts that “[p]utting hazardous materials underground increases the security of the materials.”

## **Equal Opportunity: Retrievable Disposal**

The third concern is how to act in accordance with our alleged obligations in order to minimize future burdens while at the same time not depriving people of the future of their freedom of action. NEA (1999, p. 22) states that the present generation “should not foreclose options to future generations.” This is termed the equal opportunity principle: “[i]t is of equal worth that we guarantee coming generations the same rights to integrity, ethical freedom and responsibility that we ourselves enjoy” (KASAM 1988). In other words, we should respect their freedom of action – conceived of by KASAM (1999, p. 14) as a moral value – by acknowledging that “future generations must be free to use the waste as a resource”, in view of the fact that spent fuel contains uranium and plutonium which have potential energy value. Two other factors in favor of creating the option to retrieve waste from disposal facilities are these: (1) to be able to take remedial action if the repository does not perform as expected and (2) to be able to render radiotoxic waste harmless with new technology.

Retrievability, as intended here, has to do with repositories that will be kept open for an extended period of time so that future societies have the option to retrieve the waste. One might thus argue that retrievable waste could compromise the long-term safety of any repository. However, retrievability as commonly understood in the literature implies having a temporary measure based on the assumption that at a certain point a decision will be taken to either retrieve the waste (for any purpose) or to close the repository (IAEA 2000, pp. 9–10). If one relates retrievability discussions to the question of final disposal, one can argue in favor of storage on the surface, as the “[r]etrieval of material is easier from surface facilities than from underground facilities, but geological disposal can be developed in stages so that the possibility of retrieval is retained for a long time” (IAEA 2003, p. 7).

The underlying intergenerational principles of nuclear waste management policies have been discussed in this section. We furthermore explored how a need for “the same degree of protection” for different generations has led to the conclusion that geological disposal is the most appropriate way to dispose of waste. In the following section, I shall challenge this view by reflecting on the assumptions that underlie the alleged long-term safety of geological disposal, arguing that the key weakness of this technological solution lies in the great accompanying uncertainty.

## **Policy-Making and the Principle of Diminishing Responsibility**

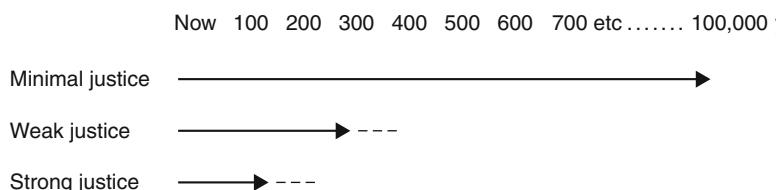
---

Elsewhere in this chapter I argued that the production of nuclear power creates certain moral obligations for the present generation to ensure future people’s well-being – in terms of resource availability – and not to harm people of the future. On the practical policy-making level, one could question to whom we owe these obligations and whether, when fulfilling these obligations, we should distinguish between people of the near and remote future

(Norton 1995). These questions date back to the early days of discussions on our obligations toward posterity. By introducing the notion of belonging to a “moral community” that shares our perception of what constitutes a good life, Martin Golding (1981, p. 62) argues that we have an obligation to produce “a desirable state of affairs for the community of the future [and] to promote conditions of good living for future generations.” However, “the more remote the members of this community are, the more problematic our obligations to them become,” Golding (1981, p. 69) states and he concludes that we should be more concerned about the more immediate generations. Daniel Callahan (1981), on the other hand, states that ignorance does not release us from our obligations when the issue at stake is whether we might harm future generations; Callahan’s notion of our negative obligations to posterity obviously extends much further than Golding’s stated positive obligations. In this section I will scrutinize the notion of reducing responsibility over the course of time, as a leading notion in policy-making, while arguing that such a distinction is inescapable when building geological repositories.

Two attempts to explore the legitimacy of a distinction between different future generations are worth mentioning, particularly as they have offered policy-making justification on nuclear waste management. The bodies to which I refer are the Swedish National Council for Nuclear Waste (KASAM) and the American National Academy of Public Administration (NAPA). The KASAM (1999, p. 27) argues along the same lines as Golding by stating that “our responsibility diminishes on a sliding scale over the course of time,” as “the uncertainties of our base of knowledge [...] of the system’s technical design, increase as a function of increasing time span perspectives.” KASAM (2005) contemplates a more extensive duty to the immediate future by introducing three principles of justice and distinguishing between the various periods of time in the future: (1) the next five generations (150 years) deserve our greatest attention; (2) the subsequent 150 years come in second place; and (3) the era beyond 300 years is considered last of all. These periods of time represent different concepts of justice in what is perceived as a “diminishing moral responsibility” perspective, as illustrated in **Fig. 12.3**.

The KASAM’s strong principle of justice states that we must ensure that our immediate descendants have “a quality of life equivalent to ours,” while the weak principle says that we need to respect and protect future people’s right to satisfy their basic needs. The minimal principle of justice, on the other hand, merely states that today’s people should not jeopardize future generations’ possibility for life. KASAM seeks justification for these proposed periods by asserting that if we define a generation as being 30 years then our imagination hardly extends beyond that of our grandchildren’s grandchildren; these five generations or 150 years are believed to represent the boundary of our “moral empathy.” Another justification that



**Fig. 12.3**

Three principles of justice with respect to the various time periods (Source: KASAM 2005, p. 440, with kind permission)

KASAM provides in connection with this principle of justice is that “our primary relationships” can barely have any influence beyond five generations. If one were to define local communities and nations in terms of “secondary relationships,” this timescale could possibly be extended to, as KASAM suggests, 300 years, beyond which predictability and positive influence “appears to be almost non-existent” (KASAM 2005).

The second initiative worth mentioning is that of the NAPA (1997, p. 7) which proposed establishing four principles of intergenerational decision-making. These principles state that we are trustees for future generations and that – according to the Brundtland’s sustainability principle – we should not deprive the future of a quality of life comparable to our own (WCED 1987). Since we know more precisely the needs and interests of our own generation and the immediately ensuing ones and since we are hardly in a position to predict the very distant future, a “rolling present” should “pass on to the next the resources and skills for a good quality of life”; NAPA (1997, p. 8) further states that “near-term concrete hazards have priority over long-term hazards that are less certain.” What remains questionable is whether the lower probability of long-term risks will make them morally less important; I will elaborate on this issue in the section [State of the Art in Technology: Challenge to Geological Disposal](#).

## Geological Disposal and Long-Term Uncertainties

---

One of the key arguments when opting for geological disposal rather than above ground storage is the alleged long-term safety issue. In safety studies it is assumed that canisters will inevitably breach and radioactive material will leak at some time in the future; canisters are therefore enclosed in engineered facilities to avoid unnecessary seepage into the environment. In addition, host earth formations are viewed as a natural barrier impeding further leakage into the biosphere, all of which should support arguments in favor of the long-term safety of repositories. This alleged safety seems, however, to be founded on a number of serious uncertainties and has therefore triggered a whole debate on whether geological disposal should not be reconsidered for the final isolation of radiological waste (this point has for instance been defended by Shrader-Frechette (1993, 1994)); more will be said about this issue in the following section.

It is curious to see how the problem of the technical unpredictability of the remote future and the associated uncertainties are addressed in long-term policy-making, particularly in relation to geological disposal. Let me illustrate this by pointing out how the radiation standards are set for the Yucca Mountains repositories in Nevada (US). According to the American Energy Policy Act of 1992, the Environmental Protection Agency (EPA)<sup>6</sup> is charged with the task of developing health and safety radiation standards for the designing of repositories.<sup>7</sup> EPA’s first proposed standards limit exposure to radiation to 15 per year, with a compliance time of 10,000 years millirem.<sup>8</sup> In objection to this it has been argued by the US National Academy of Science that “the peak risks might occur tens to thousands of years or even farther in the future” (NRC 1995); 10,000 years seemed both insufficient and arbitrary. In a successful lawsuit, the D.C. Court of Appeal subscribed to the latter conclusion ruling that the EPA had to revise these radiation standards (Vandenbosch and Vandenbosch 2007, Ch. 10). In 2005, the EPA proposed distinguishing between two future groups, i.e., the people of the next 10,000 years who could maximally be exposed to the already set 15 millirem per year standard and the period beyond 10,000 (up to one million years) for which a radiation limit of

350 millirem per year was proposed (EPA 2005). This two-tiered approach was necessary, EPA (2005, 49035) argued, in connection with long-term uncertainties which are “problematic not only because they are challenging to quantify, but also because their impact will differ depending on initial assumptions and the time at which peak dose is projected to occur.”

By applying KASAM’s Minimal Principle of Justice (of not jeopardizing the possible life of future generations) and NAPA’s preference for avoiding near-term concrete hazards in contrast to the long-term hypothetical hazards of spent fuel disposal, EPA (2005, p. 49036) concludes that “a repository must provide reasonable protection and security for the very distant future, but this may not necessarily be at levels deemed protective (and controllable) for the current or succeeding generations”. The proposed 350 millirem is the difference between the naturally occurring radiation in an average area in the USA (350 millirem per year) and that experienced in Colorado (700 millirem per year). EPA justifies this discrimination by stating that after one million years, people in Nevada will maximally experience the same level of radiation as people living in Colorado today.

On the one hand, this discrepancy in radiation standards seems understandable if one thinks that providing equal protection levels for such periods of time is virtually impossible in view of the fact that “the uncertainties for a thousand years [...] from now are large [and] they are almost incalculable when one goes to 10,000 or 100,000 years” (Kadak 1997, p. 49). On the other hand, we can question whether changing the standards for the latter reason is appropriate. As Vandenbosch and Vandenbosch (2007, p. 136) have correctly stated: “[i]f one were designing a bridge whose steel and concrete performance become more uncertain [over the course of] time, would one loosen or tighten the structural design standards if one realizes that the bridge was going to have to provide safe transport for a long period of time?” The justification provided by the EPA is furthermore believed to be flawed as these policies “threaten equal protection, ignore the needs of the most vulnerable, [and] allow many fatal exposures” of the people living in the distant future (Shrader-Frechette 2005, p. 518).

In its “final rule,” the EPA (2008) changed the radiation exposure limit for the period beyond 10,000 years to 100 millirem per year. It remains unclear what precisely motivated this change; a speculative conclusion is that that public opposition to the huge difference between the originally proposed 15 and 350 millirem for the stated periods might have triggered this adjustment in the final ruling. In June of 2008, the US Department of Energy submitted a license application to the Nuclear Regulatory Commission to build a long-term geological repository for the permanent disposal of spent fuel for a million years (DOE 2008).<sup>9</sup>

To conclude, in this section I have reviewed the underlying arguments for distinguishing between different future generations based on the low degree of predictability concerning the remote future and the fact that any positive influence on such societies is meaningless. Even though these arguments are sound, I argue that at most they provide explanations indicating why we cannot act otherwise, rather than solid moral justifications for discriminating future generations; ignorance does not release us from our temporal obligations when it comes to the question of harming future people, as correctly stated by Callahan (1981). If my arguments to the effect that the production of nuclear power creates a problem of intergenerational equity are sound, I consider the minimal principle of justice (of not jeopardizing the possibility for life) an undesirable one as it facilitates the serious discrimination of remote future generations; that is to say, we can dramatically reduce the well-being of future generations and jeopardize their health and safety without depriving them of “the possibility for life,” even though their quality of life might be much lower than ours. Whether the present level of well-being has

sufficient moral relevance to serve as a point of reference is a claim that I leave unanalyzed; I simply assume that it does; see for more information on this dispute the he debate between Wilfred Beckerman (1999) and Brian Barry (1999).

The distinction that is made in policy-making, for instance with the setting of radiation standards for Yucca Mountain repositories, seems to be rather a pragmatic solution making it possible for such repositories to remain within the margins of technical predictability for the remote future. This should urge us to reconsider our temporal moral obligations and what, in the light of recent technological developments, we ought to do with regard to future generations, assuming that “ought to” implies “can.” The following section contemplates the technological possibility of substantially reducing the waste lifetime and challenging the need for geological disposal.

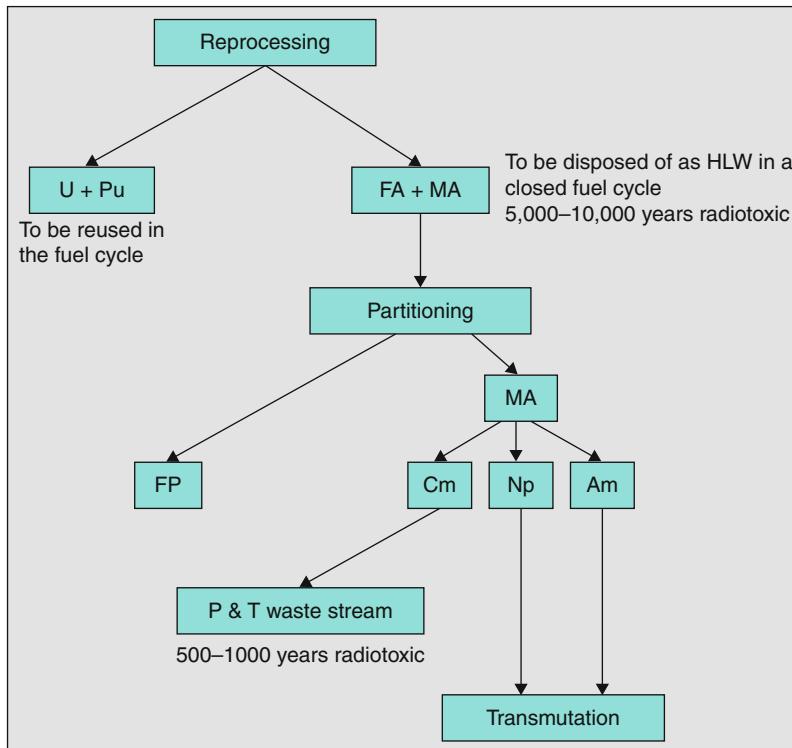
## **State of the Art in Technology: Challenge to Geological Disposal**

---

As nuclear waste is perceived to be the Achilles’ heel of nuclear energy production, serious attempts have been made to further reduce its lifetime and volume. A new technology for the latter purpose is that of Partitioning and Transmutation (P&T). Remember that spent fuel in an open fuel cycle contains uranium and plutonium, minor actinides, and fission products. Uranium and plutonium are separated during reprocessing in order to be reused; that is what amounts to a closed fuel cycle. P&T focuses on “eliminating” minor actinides, as illustrated in  Fig. 12.4. P&T complements reprocessing but does not provide an alternative solution; P&T is also referred to as an extended closed fuel cycle.

If completely successful, P&T will, it is expected, make the waste lifetime five to ten times shorter when compared to closed fuel cycle waste. After P&T, waste radiotoxicity can decay to a nonhazardous level within the space of hundreds of years (i.e., 500–1000 years). This estimated reduction in the waste lifetime is based on the assumption that all minor actinides are transmuted except for curium; the waste stream would therefore only consist of relatively short-lived fission products and curium isotopes. The latter is considered to be too hazardous to be recycled at reasonable expense and without excessive risk; curium would dominate the waste lifetime. There is a dispute about what exactly the waste lifetime will be after successful P&T. It goes beyond the scope of this work to enter into such discussions. However, for the sake of argument I adhere to the mentioned period, arguing that the scientific possibility to reduce the waste lifetime to a couple of hundred years urges us to revisit some intergenerational arguments relating to waste management. For the sake of clarity, the three different types of waste, their constituents, and the relevant waste lifetimes are all illustrated in  Table 12.1.

Some experts in the nuclear community hailed P&T in nuclear waste management but then went on to reject it for two reasons: (1) because it necessitates the building of new facilities and (2) because even after successful application some materials still remain radiotoxic (IAEA 2000; NEA-OECD 1999). Even though both arguments are sound, they do not provide sufficient grounds for rejecting P&T. In this kind of reasoning P&T is wrongly presented as an alternative to geological disposal. If my arguments in this chapter are correct, it must be asserted that P&T challenges the need for final disposal underground and places the serious alternative of repositories – for long-term storage on the surface – in a new perspective. Let me start supporting this claim by reevaluating the three main intergenerational arguments that underlie nuclear waste management policy.

**Fig. 12.4**

A schematic representation of Partitioning and Transmutation. FP stands for fission products, MA stands for minor actinides that contain curium (Cm) neptunium (Np) and americium (Am). Uranium (U) and Plutonium (Pu) are the major actinides

**Table 12.1**

Three different states of waste, the constituents, and the waste lifetimes

	Contains	Waste lifetime	Dominated by
Spent fuel	U, Pu, MA (Np, Am, Cm) + FP	$\pm 200,000$ years	Pu
HLW or vitrified waste	MA (Np, Am, Cm) + FP	$\pm 10,000$ years	Np + Am
P&T waste	MA (Cm) + FP	$\pm 500\text{--}1,000$ years	Cm

FP stands for fission products, MA stands for minor actinides that contain curium (Cm) neptunium (Np) and americium (Am). Uranium (U) and Plutonium (Pu) are the major actinides.

In objecting to above ground storage places, the IAEA (2003) draws attention to “some structural degradation of the packages and their contents [...] over time”, which makes further transfer of the waste to other storage facilities or geological repositories inevitable. The argument is that long-term safety is therefore not well served by very long periods of time in above ground storage facilities. In its recommendations in favor of geological disposal, IAEA

takes long-term safety for granted. However, the long-term safety of geological disposal depends on certain considerable uncertainties, which necessitate the sanctioning of a distinction between different future generations. If we now accept the conclusion drawn in the last section to the effect that this distinction lacks moral justification, we can argue that it would be best to avoid such uncertainties. Implementing P&T allows for the latter, as the period of necessary care for P&T waste amounts to a couple of hundred years, a period in which it is presumed that more reliable predictions can be made about a canister's status and possible seepage into the environment and whether that can reach the biosphere.

Likewise, security concerns will change. Security has to do with the unauthorized possession or theft of radiotoxic waste for the purposes of sabotage (e.g., dispersal) or proliferation. As far as sabotage is concerned, geological disposal has obvious advantages for all three above-mentioned types of radiotoxic waste as listed in ➤ *Table 12.1*: i.e., potential hazards will literally and figuratively be buried at very difficult to access depths under the ground. Hence, any sabotage concerns associated with radiotoxic waste remain evidently less in the case of geological disposal. In the case of the proliferation of nuclear weapons, however, we must distinguish between the three types of waste. Spent fuel has potential proliferation hazards, as there is still plutonium that could be separated; spent fuel might therefore best be disposed of underground (Stoll and McCombie 2001). High-level waste, however, has no potential proliferation threats, as the fissionable materials (uranium and plutonium) have already been largely extracted<sup>10</sup> and the remaining waste (minor actinides and fission products) does not lend itself to proliferation purposes. Similar reasoning is applicable to P&T waste; in other words P&T waste does not necessitate geological disposal from the avoidance of proliferation point of view.

Equal opportunity is the third intergenerational equity consideration that underlies policy-making in nuclear waste management. Nuclear waste should always be disposed of in a retrievable manner for (1) the possible future resource value of spent fuel, (2) remedial action if the repository does not operate as expected, and (3) rendering radiotoxic waste harmless with the help of new technology. By including P&T in these discussions as a technological option, one can conclude that considerations about future resource value cease to be relevant, since P&T waste comprises no potential source value in view of the fact that plutonium and the remaining uranium are separated during the earlier stage of reprocessing prior to P&T. However, retrievable disposal remains desirable in conjunction with the second and the third reasons above; i.e., even with P&T waste streams it might be necessity to adjust repositories or to render the waste harmless. This retrievability argument does not, however, support geological disposal, since retrievability is, in principle, more feasible in aboveground storage places.

P&T thus enables us to avoid the long-term uncertainties that accompany the geological disposal of long-lived spent fuel. In other words, it helps us to avoid ending up in situation in which – from a pragmatic point of view – we need to discriminate remote future generations through the way that we design and build repositories. P&T therefore puts contemporaries in a better position to fulfill the obligations emanating from the intergenerational equity problem caused by uranium depletion and, just as importantly, the longevity of nuclear waste.

Before going on to review the possible counterarguments to P&T, let us just focus on another issue, namely that of whether more predictable risk is necessarily more justifiable. As stated above, P&T helps us to avoid putting distant future generations at a disadvantage as it increases the degree of predictability. Does such increased predictability imply that the risks in the coming 500–1,000 years are justifiable? Making the relevant timescales more predictable

makes appropriate risk assessment possible. However, the questions about the acceptability of these risks for future generations and, more importantly, the additional risks posed to the present generation remain unanswered. This brings us back to the issue at the heart of this matter, namely that of intergenerational equity. Together with Kloosterman I have compared open and closed fuel cycles in terms of their conflicting moral values and have argued that the fuel cycle choice should be presented within the framework of intergenerational equity (Taebi and Kloosterman 2008). The closed fuel cycle improves uranium supply certainty and brings fewer long-term radiological risks and proliferation concerns. On the other hand, it compromises short-term public health and safety, and also security. The trade-offs inherent in opting for the fuel cycle are, as we have argued, reducible to a primary trade-off between the present and the future. If we view P&T as an extension of the closed fuel cycle – since without reprocessing the deployment of P&T is useless – and if we assume that P&T requires extra nuclear activities for the further irradiating of HLW so as to eliminate minor actinides, then more or less the same time-related conflicts will arise. There are conflicts of interest between the present generation and future generations; advocates of P&T need to justify why they are willing to accept additional risks as a result of P&T in order to reduce risks to people living in the distant future.

Let me clarify these intergenerational conflicts by focusing on one of the issues involved, namely, the economic considerations. NEA has evaluated the viability of P&T through Fast Reactors and Accelerator-Driven Systems and has concluded that a considerable amount of R&D will be required in the coming decades before utilization at industrial level can be considered (NEA-OECD 2002). In addition, more nuclear facilities will be needed for the further elimination of minor actinides. Why are these additional economic burdens upon the present generation permissible or even desirable? Precisely specifying the time-related dilemmas and trade-offs that need to be considered before P&T can be deployed should be a subject of future study; see for a preliminary discussion of this issue (Taebi and Kadak 2010).

## Four Counterarguments to the Feasibility and Desirability of P&T

So far I have argued that there are good reasons to believe that P&T is a potential future possibility and that if P&T is industrially feasible we might want to reconsider geological disposal for the final isolation of nuclear waste. There are, however, unanswered questions (or even objections) when it comes to applying this technology. The four main ones are (1) the reliance on industrially not yet proven technology for changing policy, (2) the continued need for final isolation and building repositories, (3) public resistance to additional nuclear facilities, and (4) the inappropriateness of relying on future societies to deal with our waste. These four counterarguments will be reviewed below.

At present P&T is only available in the laboratory; a considerable amount of R&D effort will be required before P&T can be industrialized (NEA-OECD 2002; IAEA 2004). The first question is whether the prospect of future technology should change current policy. Some people argue that envisaged technological progress justifies the current postponing of action as far as the building of repositories goes. KASAM (2005, Ch. 8) has evaluated the validity of postponing actions based on potential P&T possibilities as follows. On the one hand, it could be asserted from a utilitarian point of view that technologically better final disposal technology increases safety and reduces the risks for future generations. On the other hand, it is doubtful

whether that provides sufficient reason to shift the burden of finding a solution for final disposal to future generations. In addition, we need to make very explicit assumptions about the progress of technology in order to justify postponing action; this is referred to as the *technological fix* position as, for instance, defended by Beckerman and Pasek (2001). Although there has been evident technological progress during the last few centuries and although inductive reasoning forecasts its continuation, we cannot be sure that such progress is sufficient to deal with the risks we have created (Skagen Ekeli 2004) and we do not know whether future societies will be able to dispose of radiotoxic waste more appropriately than we are able to at present. More importantly, there is no guarantee of whether and, if so, when and to what extent P&T, at industrial level, will live up to the expectations it has created.

In Sweden, on the basis of P&T technology as it was in 2004, KASAM recommended that the nuclear waste program should be neither interrupted nor postponed and that the building of repositories should be continued as P&T cannot be cited as an alternative to final disposal (KASAM 2005). Nevertheless, it is believed that P&T development and future possibilities seriously call for the retrievability of waste so that future people's freedom of action to deal with such waste in a more appropriate way can be respected. The Canadian Nuclear Waste Management Organization (NWMO 2005) also finds P&T an interesting option as it reduces waste radiotoxicity and volume, but the organization has serious reservations about the economic and practical aspects of this technology and therefore maintains that it is not a desirable option for Canada.

The second problem is that P&T fails to make ultimate isolation redundant which means that some materials will remain hazardous and must therefore eventually be isolated from the environment. P&T cannot therefore entirely replace repositories that are, for instance, needed to dispose of tons of highly enriched uranium and weapon-grade plutonium emanating from dismantled nuclear warheads dating from the Cold War period. The American National Academy of Sciences acknowledged these concerns and called for the urgent implementation of a program (1) to burn this plutonium in reactors of existing types as mixed oxide fuel and (2) to mix this weapon-grade plutonium with highly enriched uranium and vitrify it for final geological disposal (NAS 1994). Likewise, the highly enriched uranium extracted from the dismantled nuclear warheads could, to some extent, be blended down to reactor-grade uranium suitable for use in a reactor. However, for the time being, there is no realistic way to fission all these materials in reactors, in spite of all the technical possibilities we have at our disposal. If we bear in mind that uranium and plutonium are the long-lived isotopes that necessitate geological disposal and if we take for granted the fact that the nuclear community is right about the appropriateness of geological repositories for the long-term disposal of long-lived isotopes, the need for geological disposal for military waste containing these materials will subsequently remain unchanged. Nonetheless, one can argue that we should consider the need for repositories that are directly related to nuclear energy production and realize that P&T will challenge this need, as fewer long-term concerns will be involved. Even though some repositories will still be needed (particularly for dismantled military material), successful P&T deployment would substantially reduce this need.

The third issue that the application of P&T might raise is that of the lack of public acceptance of more nuclear facilities. As was noted in the section [State of the Art in Technology: Challenge to Geological Disposal](#), P&T should be preceded by reprocessing (See [Fig. 12.4](#)), which is a chemical process that has, in the past, been met with considerable public opposition, for instance in Germany. It is conceivable that P&T might elicit similar objections.

In addition, the implementation of P&T necessitates the building of more fast reactors so that the troublesome actinides can be eliminated, all of which will culminate in a shorter lifetime for the remaining waste. In light of recent accidents in the Fukushima reactors in Japan, it is to be expected that the building of new nuclear facilities might meet with public outcry. However, these kinds of objections should be viewed in the light of the central issue at the heart of this chapter, namely, the matter of intergenerational equity. On the one hand, it is true that engaging in more nuclear activities will increase the risks and potential burdens for present generations, all of which might not be easily accepted by the public. On the other hand, we should realize that the already generated waste needs to be dealt with anyway. If my arguments are sound and if the current solution involving geological disposal necessitates a distinction between different future generations that is morally indefensible, then it is worth discussing different possibilities to reduce future burdens, one of which is the P&T solution. Nevertheless, the issue of public acceptance is highly relevant and will probably remain problematic.

The fourth objection to surface storage (irrespective of the waste lifetime) is that it forces us to rely on future societies for the possible further treatment and final disposal of waste (NEA-OECD 1995). It has been argued that near future geological disposal complies better with intergenerational equity, as it does not involve passing on our responsibilities to our descendants and it imposes fewer safety and security burdens on the present generation. Axel Gosseries (2008), for instance, argues that from the viewpoint of intergenerational equity, the “seriousness risk of malevolent use” calls for the early disposal of spent fuel rather than for storage.

The latter objection reveals an intergenerational conflict, not one between the interests of present and future generations (as has been outlined in this chapter), but rather one between the interests of different future generations. One can indeed defend the argument that disposing of waste now complies better with equity toward generations of the near future since, instead of passing on the responsibility of dealing with this problem, we will have taken care of it ourselves. On the other hand, assuming that geological disposal will put distant future generation at a disadvantage, it would be a good intergenerational equity argument to avoid such discrimination. The key question here is how can we rank the interests of people living during different eras in the future. If my arguments in the section [Policymaking and the Principle of Diminishing Responsibility](#) are sound and if it is the case that we ought not to treat distant future generation differently just because they happen to live in the more distant future, then it becomes less defensible for intergenerational equity to require us to dispose of the waste in geological repositories.

## Conclusions

---

Nuclear energy is one of the clearest examples of a technology that creates risks beyond the generational borders and which therefore brings about issues of intergenerational equity; because with nuclear energy production and consumption we are depleting a nonrenewable resource (uranium) that will not then be available to future generations and because, more importantly, the remaining waste needs to be isolated from the biosphere for a very long time. Principles of intergenerational equity have already been influential in nuclear waste management policies, the argument being that we should not impose *undue burdens* on future generations. In concrete terms, we have been urged to ensure future people’s safety and security whilst giving them equal opportunities. In this chapter, I have explored how these

intergenerational equity principles have been included in policy-making and especially, how they deal with the issue of long-term risks and the associated uncertainties.

The consensus within the nuclear community seems to be that burying spent fuel in geological repositories rather than keeping it in surface storage places is the best solution as repositories are believed to be safer and more secure in the long run. This long-term safety seems to be disputable, though, as it relies on great long-term uncertainties which, in turn, necessitate making a distinction between different future people. The latter distinction does, however, lack solid moral justification, which should urge us to reconsider our temporal moral obligation and what we ought to do with regard to future generations, in the light of recent technological developments. The technological possibility of substantially reducing the waste lifetime through Partitioning and Transmutation (P&T) is believed to challenge geological disposal, placing long-term surface storage in a new perspective. Deploying P&T sheds new light on the social and technical predictability capacity for the future and enables contemporaries to discharge their temporal obligations better, particularly, the obligation not to harm those living in the distant future.

In the nuclear community, P&T was wrongly presented as an alternative to geological disposal and so it was subsequently rejected as a possibility. The related misunderstanding hinged on the fact that new facilities do have to be built and the need for final isolation does remain. Even though both arguments are sound, P&T challenges the need for final disposal underground, thus placing long-term surface storage in a new perspective. Before P&T can be phased in on an industrial scale in the next few decades substantial investments will first have to be made and there is no guarantee of success, as the future development depends on very many technical, social, political, and economic factors. More importantly, the introduction of P&T creates additional risks and burdens for the present generation and serious trade-offs need to be made before acceptance and deployment can be achieved. There are also other objections to the application of P&T, such as the reliance placed on near future societies to deal with the waste and the not to be overlooked fact that it is an industrially not yet proven technology. Nonetheless, P&T helps us to avoid ending up in situation in which we – from a pragmatic point of view – need to place remote future generations at a disadvantage because of the way in which we design and build our present-day repositories. Even the potential possibility to diminish “undue burden” with respect to the future is too relevant to be neglected in discussions relating to nuclear waste management and how we perceive our moral obligations toward posterity.

---

## Further Research

Before P&T can be introduced it is, above all else, further research that is needed so that the required technology can be further developed. It is both the reprocessing technologies (multiple recycling) and the development of fast reactors that needs to be further refined. In terms of philosophy further research is required in at least the following three subject areas. Firstly, there is the issue of how to defend obligations to future generations. The matter of how to deal with the non-identity problem remains a philosophically moot point. Secondly, we need to spell out the extent of our obligations to posterity, particularly when they conflict with obligations to present generations; some thoughts on this issue could be read in the present discourse (Taebi 2011). Thirdly, further research should be conducted into the matter of how to balance intergenerational equity with intragenerational equity. In other words, what should we do

when complying with intergenerational equity creates a situation of injustice within the present generation? This problem manifests itself in the case of multinational waste repositories. In other words, from the perspective of intergenerational equity, multinational repositories are beneficial since the total number of facilities that carry a potential burden for future generations will inevitably be reduced. On the other hand, multinational repositories create the problem of intragenerational *injustice* between the participating countries, since host nations must accept the waste of other nations.

## Notes

---

1. For detailed discussions on intergenerational justice readers are referred to these articles: Gosseries (2002), Meyer (2008), and the following two collections Gosseries and Meyer (2009) and Tremmel (2006).
2. Justice, fairness, and equity are used interchangeably in the relevant literature sources. In this chapter I do not intend to go into great depth on these philosophical discussions. *Intergenerational equity or justice* are referred to here as the equitable distributions of risks and burdens across generations.
3. If the same rate of uranium consumption as that of 2008 were continued there would be enough *reasonably priced uranium* available for approximately 100 years. However, it has been forecast that many countries will join the nuclear club in the next couple of decades, all of which will substantially affect the demand for uranium; in the high-growth scenario approximately half of these resources will be depleted by 2035. It is important to note that this uranium availability constitutes a reference to its geological certainty and production costs. If we include estimations of all the available resources, this will rise substantially to thousands of years; (see IAEA-NEA 2010). Furthermore, there is another fission material available that could be used for the production of nuclear power, namely, Thorium (Th). Thorium is naturally more abundant than uranium, but it brings with it different issues such as substantial security burdens (IAEA 2005); None of the above points do, however, undermine the basic rationale that the present generation is depleting a nonrenewable resource, particularly, the currently reasonably priced supplies.
4. The term “acceptance” might be misleading here, as we are not referring to what is actually *accepted* by the public, but rather to what is taken to be the greatest allowable risk in policy-making, e.g., as in maximum exposure when setting current radiation standards.
5. I received this Figure following personal contact with Jan Leen Kloosterman from the Department of Radiation, Radionuclides and Reactors at Delft University of Technology.
6. The Environmental Protection Agency is the United States’ federal agency in charge of protecting human health by protecting the natural environment. Projects that involve affecting the natural environment – such as the disposal of nuclear waste – need to be first approved by EPA.
7. These standards must further be incorporated into licensing by the Nuclear Regulatory Commission (NRC) regulations. Compliance with these standards must then be demonstrated by the US Department of Energy (DOE). For an overview of how these organizations are connected to each other, see Vandebosch and Vandebosch (2007, Ch. 8).
8. Rem and Sievert (Sv) indicate radiation exposure in order to determine radiation protection. 1 sievert (Sv) = 100 rem.

9. The first draft of this chapter was written in 2008 when the Yucca Mountains repository was still considered to be the most feasible option for the geological disposal of American waste. However, following the most recent presidential elections in the USA, the Obama administration has excluded the Yucca Mountains site as an option; (see Josef Hebert 2009). Precisely how the American waste (that is currently stored at nuclear reactor sites) should be dealt with remains a subject of discussion. In 2010 a Blue Ribbon Commission on America's Nuclear Future started exploring the possibilities; see <http://brc.gov>.
10. It is important to notice that during reprocessing some trace amounts of plutonium will remain in the waste stream. This amount is, however, insignificant for proliferation or energy production purposes.

## Acknowledgments

---

An earlier draft of this chapter was presented at the Energy and Responsibility conference which was held in May 2008 in Knoxville, Tennessee. I wish to thank the audience of that conference and in particular Dennis Arnold. I further wish to thank Ibo van de Poel, Stephen Gardiner, Jan Leen Kloosterman, Dominic Roser, and Lara Pierpoint for their useful comments and corrections. The usual disclaimer with regard to authorial responsibility applies.

## References

---

- Athanasiou T, Baer P (2002) Dead heat: global justice and global warming. Seven Stories Press, New York
- Barry B (1999) Sustainability and intergenerational justice. In: Dobson A (ed) Fairness and futurity: essays on environmental sustainability and social justice. Oxford University Press, New York, pp 93–117
- Beckerman W (1999) Sustainable development and our obligations to future generations. In: Dobson A (ed) Fairness and futurity: essays on environmental sustainability and social justice. Oxford University Press, New York, pp 71–92
- Beckerman W, Pasek J (2001) Justice, posterity, and the environment. Oxford University Press, New York
- Bunn M (2000) The next wave: urgently needed new steps to control warheads and fissile material. Carnegie Non-Proliferation Project, Carnegie Endowment for International Peace; Harvard Project on Managing the Atom, Belfer Center for Science and International Affairs, Harvard University
- Bunn M. et al. (2001) Interim storage of spent nuclear fuel. Managing the Atom Project, Harvard University, Cambridge, MA, and Project on Sociotechnics of Nuclear Energy, University of Tokyo
- Callahan D (1981) What obligations do we have to future generations? In: Partridge E (ed) Responsibilities to future generations: environmental ethics. Prometheus Books, Buffalo, pp 73–85
- Caney S (2006) Cosmopolitan justice, responsibility, and global climate change. *Leiden J Int Law* 18(04):747–775
- DOE (1997) Nonproliferation and arms control assessment of weapons-useable material storage and excess plutonium disposition alternatives. Washington, DC: U.S. Department of Energy
- DOE (2008) Yucca mountain repository license application. Washington, DC: Department of Energy, Licensing Support Network, DEN001592183-MOL.20080501.0021
- EPA (2005) Public health and environmental radiation protection standards for yucca mountain. 40 CFR Part 197, Part II. Environmental Protection Agency
- EPA (2008) Public health and environmental radiation protection standards for Yucca Mountain; Final Rule. 40 CFR Part 197, Part III. Washington, DC: Environmental Protection Agency
- Feinberg J (1981) The right of animals and unborn generations. In: Partridge E (ed) Responsibilities to future generations: environmental ethics. Prometheus Books, Buffalo, pp 139–150
- Gardiner SM (2001) The real tragedy of the commons. *Philos Public Aff* 30(4):387–416

- Gardiner SM (2003) The pure intergenerational problem. *Monist* 86(3):481–501
- Gardiner SM (2004) Ethics and global climate change. *Ethics* 114(3):555–600
- Gardiner SM (2006a) A perfect moral storm: climate change, intergenerational ethics and the problem of moral corruption. *Environ Values* 15(3):397–413
- Gardiner SM (2006b) Why do future generations need protection? *Cahier de la Chaire Développement durable EDF-École polytechnique*: 1–16
- Golding MP (1981) Obligation to future generations. In: Partridge E (ed) *Responsibilities to future generations: environmental ethics*. Prometheus Books, Buffalo, New York, pp 61–72
- Gosseries A (2002) Intergenerational justice. In: LaFollette H (ed) *The Oxford handbook of practical ethics*. Oxford University Press, Oxford, pp 459–484
- Gosseries A (2008) Radiological Protection and Intergenerational Justice. In: Eggermont G, Feltz B (eds) *Ethics and radiological protection*. Academia-Bruylant, Louvain-la-Neuve, pp 167–195
- Gosseries A, Meyer L (eds) (2009) *Intergenerational justice*. Oxford University Press, Oxford
- IAEA (1995) The principles of radioactive waste management. In: *Radioactive waste safety standards programme*. IAEA, Vienna
- IAEA (1998) Interim storage of radioactive waste packages. In: Technical reports series, No 309. IAEA, Vienna
- IAEA (2000) Radioactive waste management – turning options into solutions. IAEA 3rd Scientific Forum, Vienna
- IAEA (2003) The long term storage of radioactive waste: safety and sustainability. In: *A Position Paper of International Experts*, IAEA, Vienna
- IAEA (2004) Technical implications of partitioning and transmutation in radioactive waste management. IAEA, Vienna
- IAEA (2005) Thorium fuel cycle – Potential benefits and challenges. IAEA, Austria
- IAEA (2007) IAEA safety glossary, terminology used in nuclear safety and radiation protection. IAEA, Vienna
- IAEA-NEA (2010) Uranium 2009: resources, production and demand. In: A joint report by the OECD Nuclear Energy Agency and the International Atomic and Energy Agency. IAEA and NEA-OECD, Paris
- Josef Hebert H (2009) Nuclear waste won't be going to Nevada's Yucca Mountain, Obama official says. *Chicago Tribune*, March 6
- Kadak AC (1997) An intergenerational approach to high-level waste disposal. *Nucl News* 40:49–51
- KASAM (1988) Ethical aspects on nuclear waste, in SKN Report 29. National Council for Nuclear Waste (KASAM), Stockholm
- KASAM (1999) Nuclear waste – State-of-the-art reports 1998. National Council for Nuclear Waste (KASAM), Stockholm
- KASAM (2005) Nuclear waste state-of-the-art reports 2004. National Council for Nuclear Waste (KASAM), Stockholm
- Macklin R (1981) Can future generations correctly be said to have rights? In: Partridge E (ed) *Responsibilities to future generations: environmental ethics*. Prometheus Books, Buffalo, pp 151–156
- Meyer LH (2008) Intergenerational justice. In: Zalta EN (ed) *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/justice-intergenerational/>
- Meyer LH, Roser D (2006) Distributive justice and climate change. The allocation of emission rights. *Analyse Kritik* 28:241–267
- NAPA (1997) Deciding for the future: balancing risks, costs, and benefits fairly across generations. A report for the U.S. Department of Energy. National Academy of Public Administration, USA
- NAS (1994) Management and disposition of excess weapons plutonium. Report of the National Academy of Sciences. The National Academy of Sciences, Committee on International Security and Arms Control
- NEA-OECD (1984) Long-term radiation protection objectives for radioactive waste disposal. Report of a group of experts jointly sponsored by the Radioactive Waste Management Committee and the Committee on Radiation Protection and Public Health. Nuclear Energy Agency, Organisation for Economic Co-operation and Development, Paris
- NEA-OECD (1995) The environmental and ethical basis of geological disposal of long-lived radioactive wastes: a collective opinion of the radioactive waste management committee of the nuclear energy agency. Nuclear Energy Agency, Organisation for Economic Co-operation and Development, Paris
- NEA-OECD (1999) Progress towards geologic disposal of radioactive waste: where do we stand? An international assessment. Nuclear Energy Agency, Organisation for Economic Co-operation and Development, Paris
- NEA-OECD (2002) Accelerator-driven systems (ADS) and fast reactors (FR) in advanced nuclear fuel cycles: a comparative study. Nuclear Energy Agency, Organisation for Economic Co-operation and Development, Paris
- Norton B (1995) Future generations, obligations to. In: Reich WT (ed) *Encyclopedia of bioethics*, 2nd edn. Macmillan, New York, pp 892–899
- NRC (1966) *Understanding risk: informing decisions in a democratic society*. National Research Council (NRC), National Academy Press, Washington DC

- NRC (1995) Technical basis for Yucca Mountains standards. Committee on the Technical Basis for Yucca Mountains Standards, Board on Radioactive Waste Management, National Research Council, Washington, DC
- NRC (1996) Nuclear wastes: technologies for separations and transmutation. Committee on Separations Technology and Transmutation Systems, Board on Radioactive Waste Management, National Research Council (NRC), Washington, DC
- NWMO (2005) Choosing a way forward; the future management of Canada's used nuclear fuel (final study). Nuclear Waste Management Organization, Ottawa (Ontario), Canada
- Page E (1999) Intergenerational justice and climate change. *Polit Stud* 47(1):53–66
- Page E (2006) Climate change, justice and future generations. Edward Elgar Publishing, Cheltenham, UK
- Parfit D (1983) Energy policy and the further future: the identity problem. In: MacLean D, Brown PG (eds) *Energy and the future*. Rowman and Littlefield, Totowa, pp 166–179
- Parfit D (1984) Reasons and persons, 1987th edn. Clarendon, Oxford
- Rawls J (1971) A theory of justice, revised edition. 1999 edn. The Belknap Press of Harvard University Press, Cambridge, MA
- Raz J (1986) Right-based moralities. In: Waldron J (ed) *Theories of rights*. Oxford University Press, Oxford, pp 182–200
- Schwartz T (1978) Obligations to posterity. In: Barry B, Sikora I (eds) *Obligations to future generations*. Temple University Press, Philadelphia, pp 3–13
- Shrader-Frechette K (1993) Burying uncertainty: risk and the case against geological disposal of nuclear waste. University of California Press, Berkeley
- Shrader-Frechette K (1994) Equity and nuclear waste disposal. *J Agric Environ Ethics* 7(2):133–156
- Shrader-Frechette K (2005) Mortgaging the future: dumping ethics with nuclear waste. *Sci Eng Ethics* 11(4):518–520
- Shue H (2003) Climate change. In: Jamieson D (ed) *A companion to environmental philosophy*. Blackwell, Malden, pp 449–459
- Skagen Ekel K (2004) Environmental risks, uncertainty and intergenerational ethics. *Environ Values* 13(4):421–448
- Stoll R, McCombie C (2001) The role of geologic disposal in preventing nuclear proliferation. Paper read at Proceedings of the 9th international high-level radioactive waste management (IHLRWM) conference, Las Vegas
- Taebi (2011) The morally desirable option for nuclear power production. *Philos Technol* 24(2):169–192
- Taebi B, Kadak AC (2010) Intergenerational considerations affecting the future of nuclear power: equity as a framework for assessing fuel cycles. *Risk Anal* 30(9):1341–1362
- Taebi B, Kloosterman JL (2008) To recycle or not to recycle? An intergenerational approach to nuclear fuel cycles. *Sci Eng Ethics* 14(2):177–200
- Tremmel JC (ed) (2006) *Handbook of intergenerational justice*. Edward Elgar Publishing, Cheltenham/Northampton
- Vandenbosch R, Vandenbosch SE (2007) *Nuclear waste stalemate: political and scientific controversies*, vol 61. The University of Utah Press, Salt Lake City
- WCED (1987) *Our common future*. World Commission on Environment and Development (WCED), Oxford
- Wilson PD (1999) *The nuclear fuel cycle; from ore to wastes*, 2nd edn. Oxford University Press, Oxford
- WNA. *World Nuclear Power Reactors & Uranium Requirements 2010*, Information Paper (2 March 2010). World Nuclear Association (WNA) 2011. <http://www.world-nuclear.org/info/reactors.html>

# 13 Climate Change as Risk?

Rafaela Hillerbrand

RWTH Aachen University, Aachen, Germany

<i>Introduction</i> .....	320
<i>Climate Science or Climate Fiction?</i> .....	322
Human Influence on the Climate .....	322
Impacts of Climate Conditions on Humans .....	324
A Short Plea for an Anthropocentric Approach .....	326
<i>Predicting Complex Systems</i> .....	326
<i>Knowing-How in Scientific Modeling</i> .....	328
Epistemic Uncertainties .....	328
Reducing Uncertainty .....	330
<i>Limits of Common Decision Approaches When Applied to Climate Change</i> .....	331
Interdependencies Between Moral and Epistemic Issues .....	331
Expected Utility Maximization and the Precautionary Principle .....	333
<i>Summary and Further Research</i> .....	336

**Abstract:** This chapter analyzes the types of uncertainties involved in climate modeling. It is shown that as regards climate change, unquantified uncertainties can neither be ignored in decision making nor be reduced to quantified ones by assigning subjective probabilities. This poses central problems as therefore the well-known elementary as well as probabilistic decision approaches are not applicable. While a maximized-utility approach has to presuppose probability estimates that are not at hand for climate predictions, the precautionary principle is not capable of adequately implementing questions of fairness between different nations or generations. Thus an adequate response to global warming must deal with an intricate interplay between epistemic and ethical considerations. The contribution argues that the epistemic problems involved in modeling the climate system are generic for modeling complex systems. Possible paths for future research to circumvene these problems are adumbrated.

## Introduction

---

Human beings are part of the biosphere and as such always have and always will impact the Earth's ecosystem. Due to the large number of humans living presently and due to technological changes, this impact is greater than ever. But, for the purpose of this handbook more important, unlike past generations we are well aware of our impact on the ecosystem Earth. And even more so: Science provides us with sophisticated tools that estimate this impact and tells us, for example, that the climate system is particularly vulnerable to present anthropogenic impact.

Though it is moral considerations that make us reason about global warming, determining the adequate response to this threat relies on findings from the empirical sciences. Assessing, for example, the adequate greenhouse gas reduction scheme, the pros and cons of cap and trade or carbon tax, or even the very question as to how much we have to cut down the emissions hinges on empirical prognoses. Decisions are to be based on the best knowledge available. Most scientists agree that, at least for the time being, unquantified uncertainties are inevitably connected to predictions of climate models. This, however, does not imply that these predictions are unscientific. Nor do uncertainties justify political inaction. For climate change, just like for many environmental problems, the best knowledge available today is provided by science (e.g., Oreskes 2004). This knowledge comprises not only the prognoses themselves – prognoses on, say, the increase of global mean temperature, sea level rise, or the effects all these changes have on the fishing industry. Our best scientific knowledge available today is the prognoses *plus* information on their reliability (Hillerbrand 2010). Increasing the reliability commonly coincides with reducing the uncertainties of the predictions – an enterprise of major significance within climatology. However, uncertainties remain and will remain. When practical reasoning is based on scientific facts, these uncertainties have to be considered, one way or the other.

Though uncertainty is nothing peculiar to climate modeling or even scientific forecasts, uncertainties arising here seem to differ from uncertainties associated with everyday prognoses. We expect science to provide us with quantitative information – even on the uncertainty of its predictions. And indeed, this is particularly what climatology provides us with. The Intergovernmental Panel on Climate Change (IPCC), for example, predicts that a doubling in carbon dioxide concentration “is likely to be in the range 2°C to 4.5°C with a best estimate of

about  $3^{\circ}\text{C}$ , and is very unlikely to be less than  $1.5^{\circ}\text{C}$ . Values substantially higher than  $4.5^{\circ}\text{C}$  cannot be excluded, [...]” (IPCC 2007, p. 12). Expressions like *likely* or *very unlikely* are thereby interpreted as probability statements: *Likely* corresponds to probabilities higher than 66%, while *very unlikely* denotes probabilities smaller than 10%.

In the first part of this chapter, the climate and impact models and, if applicable, the probabilistic statements they provide are analyzed. It is argued that, despite to all appearances, uncertainties arising in quantitative scientific predication cannot be fully quantified. Section [● Climate Science or Climate Fiction?](#) addresses the question as to what is the very thing that is uncertain about present climate predictions. Not aiming at an exhaustive overview, this section highlights difficulties in modeling the climate system and climatic impacts on humans. On epistemic grounds, two types of uncertainty are distinguished, namely, *parameter* and *conceptualization uncertainty*. In section [● Predicting Complex Systems](#), problems in predicting the impact of anthropogenic greenhouse gas emissions are related to more general problems concerned with complex systems. There exists a widespread and appropriate scientific specification of the demotic notion of a complex system. It defines a complex system as one with more than three degrees of freedom which are coupled to each other via feedbacks. Mathematically, this is expressed by nonlinear evolution equations. This notion must presuppose a model of the target system, that is already on such a high level of abstraction that the notion of a *degree of freedom* makes sense. As this is a chapter mainly on epistemic aspects, it does refrain from this technical notion and adopts Shackel’s definition of a complex system throughout the paper. Shackel’s notion is completely in epistemic terms and allows to picture the climate system as complex with respect to some features, while simple with respect to others. Section [● Knowing-How in Scientific Modeling](#) contends that in modeling complex systems, non-propositional knowledge, that is, competencies or abilities, as opposed to propositional knowledge become of great importance. Unlike the latter, these competencies may only be learned by practice – by working within a certain scientific field. What is good practice cannot be captured in explicit definitions or in numerical figures assessing the reliability of the practice. Assigning numerical figures, for example, in the form of probabilities, to the reliability of models and the conceptualization uncertainty associated to their predictions thus is (at least in parts) misleading. It is shown that as regards climate change, unquantified uncertainties can neither be ignored in decision making nor be reduced to quantified ones by assigning subjective probabilities. As argued in section [● Limits of Common Decision Approaches When Applied to Climate Change](#), this epistemic feature of the climate system is of central relevance for practical reasoning as unquantified uncertainties render the applicability of standard decision approaches impossible. Particularly, the uncertainties with which decision making based on these prognoses has to deal with comprise but go well beyond what scientists commonly refer to as uncertainty (i.e., the variance or width of some probability distribution). On the basis of the distinction between risk, uncertainty and ignorance as known in technology assessment, the impact of greenhouse gas emissions are identified as a genuine situation under uncertainty because the reliability of the models involved cannot be fully quantified. It is asked as to how classical decision approaches like expected utility maximization as well as a precautionary approach may deal with the uncertainties specific for climate modeling. Section [● Summary and Further Research](#) provides a summary resulting in an outlook on future research that may provide a way of how to adequately incorporate unquantified uncertainties into practical decision making.

## Climate Science or Climate Fiction?

---

Prognoses on the future are always uncertain, and so are prognoses on the future climate. So what exactly does uncertainty mean as regards global warming? Before sketching some features of models relevant for the discussion of uncertainties in this chapter, a word of caution. This contribution does not aim at an exhaustive overview on virtues and vices of present climate modeling, hence the analysis is, willy-nilly, biased. This shall, however, not distract from the fact that, despite all the shortcomings, these are scientific models and the best ones available. Present climate models are readily able to well reproduce the natural climate variability (e.g., Houghton 1995). Predictions obtained from various models widely agree with one another, the connection between carbon dioxide concentration and global average temperature seems to be well supported not only by numerical simulations, but also by measurements in ice cores (e.g., Wilson et al. 2000).

## Human Influence on the Climate

---

Not only the emission of greenhouse gases like carbon dioxide, emitted when burning fossils fuels, but various human activities impact and have always impacted on the climate: Building development, slash and burn, farming, the regulation of inland waters, etc., change the earth's surface and thus the amount of radiation backscattered from the earth into the universe as well as ground-level atmospheric currents (e.g., Mackenzie et al. 2001, pp. 51–82; Wilson et al. 2000, 240ff.). This in turn impinges on the atmospheric mean temperature on local and, in parts, on a global scale. The settling of man some thousand years ago, for example, and the corresponding crossover from nomadism to farming was accompanied by vast forest clearing and thus had significant and sustainable impact on the climate.

Already in as early as 1896, S.A. Arrhenius predicted a climatic change due to anthropogenic emissions of carbon dioxide in the wake of the industrial revolution. Today, we know not only that carbon dioxide is a greenhouse gas (i.e., it absorbs and emits radiation within the thermal range), but also that its atmospheric concentration increased from 280 ppm (parts per million) before the industrial revolution to 379 ppm in 2005. What we do not fully understand, however, is as to how the increase in atmospheric greenhouse gas alters the global mean temperature. Due to feedbacks between various components of the climate system, which comprises not only the atmosphere, but also bio-, hydro-, litho-, and cryosphere, it is not isolated cause-and-effect chains that determine the state of the climate system. An initial warming of the atmosphere caused by an increase in greenhouse gas concentration may be significantly enhanced or reduced. Higher mean temperature enhances, for example, the growth of oceanic phytoplankton (cp. Mackenzie et al. 2001, pp. 51–82; Wilson et al. 2000, 240ff.). On dying, this phytoplankton produces cloud condensation nuclei. More phytoplankton thus may increase the backscattering ratio of the sky cover and hence counteract the initial temperature rise (e.g., Idso 2001, p. 325). At the same time, increasing the atmospheric temperature reduces the level of the permafrost soils in the tundras of Siberia and Canada; huge amounts of the greenhouse gases carbon dioxide and methane that are stored in these permafrost soils are released and thus reinforce the warming of the atmosphere.

The state of the climate system results from a complex interplay between various factors. Due to feedbacks, the future course of the climate system cannot simply be determined from an extrapolation of its past or present state. Climate models, aiming to represent the major causal mechanisms in the climate system and investigated numerically, are the only way today to gain insight into the future climate. Not all feedbacks are known or understood in every detail; the resolution of some feedbacks requires too high spatial and temporal resolution to be achieved with today's computational power. Climate predictions, like those reported in the IPCC reports are commonly based on so-called general circulation models (GCM). Here, only atmosphere and hydrosphere are dynamically simulated, coupled via material and energy exchange; all other spheres of the climate system – bio-, litho-, and cryosphere – are only incorporated as static boundary conditions (cp. Betz 2009a, b). Hence major feedbacks in the climate system are not fully resolved in the models, for example, the interchange of carbon dioxide between plants and atmosphere. But also purely atmospheric processes may not be adequately represented: Following the IPCC reports, uncertainty in present climate predictions is mainly due to a lack of understanding of the radiative properties of clouds.

The sketched complexity of the climate system with its numerous components that are linked to each other via feedbacks and the intricate nature of the processes induce that not all causal mechanisms are captured by climate models. This renders their predictions uncertain. This type of uncertainty is referred to as *model conceptualization uncertainty* (Hillerbrand 2010). Note that this terminology does not distinguish between sources of uncertainty and the uncertainty itself; however, this equivocation does not seem troubling. Next to the model conceptualization uncertainty, there is yet another source of uncertainty, referred to as *parameter uncertainty* here. Climate models require input concerning, for example, the future level of greenhouse gas or aerosol concentrations; the numerical value of these quantities hinges on factors like the growth of world population, economic growth, and the course of future energy or social policy. The term "parameter uncertainty" is broadly understood here, encompassing what, for example, Refsgaard et al. (2006) refer to as uncertainty due to model input and model parameters (more narrowly understood). Apart from the twofold distinction and the threefold distinction by Refsgaard et al. (2006), various authors distinguish also uncertainties arising due to model context, model assumptions, expert judgment, or indicator choice (cp. van der Sluijs 1997; van der Sluijs et al. 2003; Walker et al. 2003). These and other more detailed distinctions prove very useful for practice; however, from the epistemic point of view pursued in this chapter, there is no need to distinguish model parameters and model input.

So-called energy scenarios assess, amongst others, rates of future anthropogenic release of greenhouse gases and aerosols and thus provide the input for climate models. The relation between the release of gases and political and economic developments is rather intricate. The second largest anthropogenic source of atmospheric carbon dioxide, for example, is land use – an area very sensitive to political and economic decisions (e.g., Houghton 1995). Also of great importance is the efficiency of future energy conversion systems or the energy intensity of industries – factors determined by future technological developments which are hard to predict. The input parameters provided by the energy scenarios are, and remain, highly uncertain. This uncertainty translates to the uncertainty of the climate models that take these parameters as input. The IPCC (and others) reacts to this uncertainty and invokes the term *climate projections*, instead of *predictions* in order to highlight the dependency of the output on the considered energy scenario.

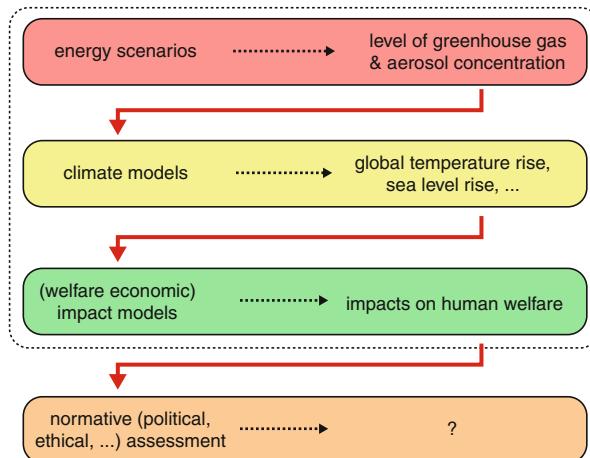
## Impacts of Climate Conditions on Humans

---

Model conceptualization and parameter uncertainty render it impossible to predict with certainty the exact impact anthropogenic actions have on the future climate. Likewise, the precise effect climatic changes exert on humans can never be predicted with certainty. That these influences may be very severe, however, does not stand to reason. For example, the European revolutions of 1789 and 1848 – though a result of the longstanding social, political, and economic disparities – were initiated after years of bad weather, bad crops, and high corn prices. Not only local, but also global weather phenomena had known impact on societies. In the tenth century, for example, worldwide drought may have wiped out the Tang Dynasty in China and that of the Mayan civilization in Mexico (Yancheva et al. 2007). Only minute changes in average climate conditions may significantly impact human well-being. A drop of average annual global temperature by, say, only 1°C may shorten the vegetation period near the polar circle, that is, in Canada, Finland and Iceland, by 3–4 weeks. Note in particular that, despite the large variations in the atmospheric temperature throughout Earth's history, the prognosticated temperature rise is much faster than all changes currently known within the last 10,000 years. The presently predicted change in global mean temperature leads in most assessments to overall negative consequences, which are most severe in those countries that are already the worst off today.

The impact, which changes in mean temperature have on humans, hinges on many factors, ranging from the way of farming to the height of landmass above sea level and people's capability to adapt. Major global warming impacts are expected via rising sea level, or increased frequency and intensity of extreme weather events such as floods, extreme droughts, heavy storms, or rainfalls. The expected changes in these factors are commonly provided by climate models. This information then serves as input to so-called impact models that estimate the influence of global warming and its geological consequences on human well-being. Hereby, well-being is understood very broadly, synonymous with welfare or a good life; a more detailed understanding is not required for the purpose of this chapter. Most often, impact models are welfare economic models estimating the impact of global warming, related changes in weather, etc., on a wide range of market and nonmarket sectors (e.g., Tol 2002; Nordhaus and Boyer 2000; Solomon et al. 2007; Stern 2007; Nordhaus 2008). Note that, unless one assigns an intrinsic value to the climate system (or some of its components), the practical debate on how to react to the threats of global warming must be based on the predictions of these impacts models: Within an anthropo- or pathocentric approach, a mere rise in the global atmospheric mean temperature is not per se morally wrong or even morally relevant; climate change only becomes a moral problem because of its impact on the well-being of human (or other sentient) beings (cp. Hillerbrand 2009, 2010; Hillerbrand and Ghil 2008). It is important to note that also non-consequentialists may agree with this claim. The linkage of climate and impact models yields an uncertainty cascade as depicted in  Fig. 13.1.

The information provided by the energy scenarios supplies the input for the climate models which in turn serves as input for the impact models. Here the uncertainties associated with the modeling on all three levels add up. It is the output of this impact assessment level that provides the input for the practical evaluation of how to react to climate change. Hence even if we had perfect understanding of the climate system, we may not know the effects this change of mean

**Fig. 13.1**

**Estimating the impacts of anthropogenic emissions of greenhouse gases on the living conditions of future generations.** The *straight arrows* (broken arrow, black) correspond to “yields the output,” while *long oblique (heavy arrow, red)* arrows correspond to “is input for”. The *dashed rectangle* indicates the combination of scientific prognoses that, as a whole, serve as an input for a normative (political or moral) evaluation (c.p. Hillerbrand and Ghil 2008)

temperature, say, has on the well-being of future generations which we need to know for a moral or political assessment based on an anthropocentric approach.

Model conceptualization uncertainty is unavoidable high for impact models as they model socioeconomic processes on large temporal and spacial scales (cp. Stern 2007). This welfare-economic modeling is exacerbated as major global warming damage is expected in nonmarket sectors; the models, however, commonly monetarize all losses. For a notable exception, see Lumer (2002). This leaves many losses unconsidered or only partly considered as money has properties that nonmonetary losses do not have or have only very rudimentarily: Money may be lent and exchanged, bear interest, etc., monetary losses may be (partially) compensated by investments, . . . (Lumer 2002). Not so the loss of habitat, friends, or relatives in, for example, extreme weather events. Note that the classification depicted in Fig. 13.1 is ideal typical nature. Often impact and climate models mix (e.g., Nordhaus 2008). Moreover, in impact models, epistemic and moral values mix which further adds to the model conceptualization uncertainty. Such a (partial) mixing of impact modeling and normative assessment cannot be avoided completely: Normative assessment is needed for determining which aspects of human life are worth or necessary to consider. More problematic is the fact that modeling assumptions like the discounting rate for nonmonetary losses have to be discussed (also) on moral grounds (cp. Stern 2007). Merging of a normative and a descriptive assessment, though impossible to fully avoid, however blurs many (normative) assumptions and may render the evaluation rather opaque.

## A Short Plea for an Anthropocentric Approach

---

Though this contribution focuses on epistemic aspects, I want to dwell a little more on the presupposition of an anthropocentric ethical framework that values nature or its parts only as far as they provide some value for human (present or future) well-being. This is in sharp contrast with many popular scientific publications that seem to presuppose a clear moral obligation to preserve the climate system or (nonhuman) parts of it as such. If there is a moral obligation to preserve the climate in its present state, where does it stem from? In approaching climate-change issues from an anthropocentric moral point of view, it is only the output on the level of impact modeling and the associated uncertainties that are of relevance: There are no *a priori* obligations toward nature or its parts, and any action has to be evaluated as to how it promotes overall human welfare. Various environmentalists have criticized valuing the environment solely as a basic resource for humanity, as discussed in the present paper. Movements like “deep ecology” (Devall and Sessions 1985) or “land ethics” (Thoreau 2004) recently attracted considerable attention in environmental discussions. Their positions are genuinely non-anthropocentric: either nature as a whole or parts of (nonhuman) nature are assigned some moral value. Hence the whole ecosystems or even the climate system have to be valued for their own sake, that is, not merely due to their value for a sentient being. Note that though this avenue is not pursued in this contribution, the welfare-based approach can be generalized to other sentient beings in a straightforward manner.

Most of the proposed non-anthropocentric approaches in the literature have difficulties in dealing with moral dilemma (Krebs 1999). This is not a crucial shortcoming of such approaches, though, as a hierarchical value structure could solve this problem. A key shortcoming of non-anthropocentric approaches, however, is that they contradict Occam’s razor: a larger number of premises are needed in arguing for physiocentrism or holism, and these added premises cannot be justified any further (cp. UNFCCC 1992). Keeping the number of such metaphysical assumptions as low as possible is, however, particularly important within environmental ethics. In order to become effective, norms that, for example, rule the emissions of greenhouse gases have to be implemented on a global scale and by future generations as well. The metaphysical background shared by different cultures – or, within one culture, over several generations – seems rather limited. The assumptions of a welfare-based approach are the most likely to be shared by people from different cultural backgrounds. As impact models consider other sentient beings, if at all, only as of instrumental value, in the following, well-being refers to human well-being only.

## Predicting Complex Systems

---

Feedbacks between various subcomponents, which manifest in nonlinear evolution equations of the modeled quantities, are characteristic for the climate and socioeconomic systems. In the language of the mathematical sciences, such systems are referred to as *complex systems*. To be more precise, a complex system is one with more than three interacting degrees of freedom. Systems that are complex in this way may exhibit chaotic behavior and thus, though deterministic in principle, cannot be predicted in every detail: Very similar causes may not lead to similar effects.

Though both the climate system and socioeconomic systems are complex in the mathematical sense, this notion is not adopted here. This chapter is on epistemic issues, therefore Shackel's epistemic definition of a complex system is deployed – a nontechnical definition given purely in epistemic terms (June 2008, private communication). Here, a complex system is defined as a non-simple one. For simple systems, knowing one or reasonably many features gives reliable information of another feature (or some other features), such as the one in which you are interested. For complex systems, this relation does not hold. A system therefore may be complex either because (1) only knowing many features gives reliable knowledge about the feature of interest or (2) the relation between the feature of interest and others is unclear. Note that the two ways in which systems are complex distinguished above comprise Kuhlmann's (2010) distinction between compositionally and dynamically complex systems, respectively. However, the terminology above is broader than Kuhlmann's as he follows the mathematical modeling.

As regards information about future global warming and its impacts, the climate and socioeconomic systems are complex in the epistemic sense: Only knowledge of many features gives insight into the features of interest, and the interdependence of various features is unclear. Note that only one reason for this is that the system is complex in the mathematical sense. The two ways in which a system may be complex directly link to the (sources of) uncertainties that were introduced in section [Climate Science or Climate Fiction?](#), parameter and conceptualization uncertainty. First, not all of the numerous parameters in the models may be known with the required precision. Second, as only a complicated interplay between its numerous features determines the state of the climate system of interest here, not all relevant causal mechanisms may be adequately incorporated into the model. This may be due to a lack of knowledge or due to contingent limitations like finite computational power.

Complexity and simplicity are, in the definition adopted here, purely epistemic features. They not only depend on the modeled system, but also on which feature of the system one is actually interested in, the desired accuracy of the feature of interest, the available background knowledge, etc. Unlike the mathematical one, the epistemic notion of a complex system is neutral with respect to whether a mathematical or less rigorous description of the system is pursued. It also encompasses a less strict usage of the term that follows its demotic meaning (e.g., Parker 2006). The adopted definition of a complex system is vague, both as regards to what counts as *reasonably many* and *reliable*; there are borderline cases. It is an epistemic notion and systems which are complex as regards one feature, may be simple as regards another. Hence the very same (ontological) system may be (epistemically) complex or simple depending on the feature of interest or the required reliability. Strictly speaking, the model may represent some (real) target system as complex or simple; however, like for the mathematical use of the term, complexity is applied to systems, keeping in mind the problems associated with models and representation (e.g., Giere 2004). Note that likewise, the classification of parameter and conceptualization uncertainty is always relative to some model focusing on certain features of interest.

Consider the example of the orbital motion of the earth around the sun. This is a simple system as regards its approximate period; it is complex when one is interested in more precise information, the deviation from the ellipsoidal motion, etc. Likewise, the climate system is simple with respect to some features, and complex with respect to others. It is simple as regards the question as to whether human activity impacts the climate, but also simple with respect to the question which countries will be most harmed by climate change. Determining that it

is the less and the least developed countries that, on average, will suffer most from climate change does not invoke running nonlinear models; a close inspection of, for example, the sensitive behavior of crop yield on aridity and the limited means of adapting to such changes suffices. Following the epistemic notion of complex systems introduced above, answering the very question as to why climate change is of moral significance thus does not invoke predicting complex systems. The question, however, as to how exactly we have to react (if at all) to the global warming threat, is a question that relies heavily on more detailed information of atmospheric temperature, say, and thus on information obtained from the analysis of a complex system.

## Knowing-How in Scientific Modeling

---

### Epistemic Uncertainties

---

By specifying a range for the uncertain parameters and assigning corresponding likelihoods, the uncertainty in the modeled quantity may be expressed by the width of the corresponding probability distribution and thus by a numerical figure. This procedure is known within nonlinear science as sensitivity analysis. If the range of the height of future emissions, say, is varied within a certain interval, one may specify a parameter range for the mean atmospheric temperature. When climatologists refer to “reducing uncertainties,” they aim at curtailing the width of the probability distribution. Model conceptualization uncertainty is often treated in a similar fashion to parameter uncertainty: By varying not only the input parameters, but at the same time also varying the underlying model, the IPCC, for example, is able to specify an interval for, say, the expected range of temperature increase (e.g., IPCC 2007, p. 12). However, as shown in the following, model conceptualization uncertainty may not satisfactorily be treated in the same way as parameter uncertainty is and thus uncertainties cannot be fully quantified. There have been several attempts recently to quantify conceptualization uncertainties (e.g., Moss and Schneider 2000), but these all remain rather fragmentary.

Scientific predictions are derived from some model representing the target system of interest. Particularly in climatology, these models are implemented numerically which may cause additional complications which are, however, not considered in this chapter. Uncertainties capture the reliability of model predictions and are thus related to the reliability of the models, that is, their structure and the numerical values of their parameters. In the following, it will be shown that as regards complex systems, the reliability scientists may assign to their models is only apprehended within the community of scientists working in the field under consideration. When communicating to outsiders, essential information gets lost; hence capturing the reliability in quantitative figures is highly misleading.

The argument builds on two premises that are argued in more detail below: (1) Just like any model, climate or impact models cannot be derived from theories in a straightforward way. (2) There is no exhaustive propositional account of this derivation. For complex systems, deriving a model involves centrally what Ryle referred to as *knowing-how*, an ability or disposition, and not merely *knowledge that*, a relation between some thinker and a (true) proposition. Before defending the two premises, let me briefly dilate on the usage of the term model and theory in this contribution. Theories here are not only understood as well-developed theoretical bodies like Newtonian mechanics, but also as theoretical background

assumptions that form common beliefs within a field, for example, the efficient market hypotheses. In the semantic interpretation of theories, the relation between model and theory is well defined; its notion of model, however, is a great distance away from the usage of the term *model* within climatology or economy. Deriving a model from a theory invokes more than merely setting numerical values for some free parameters. This chapter thus follows the more recent accounts of models used by Cartwright, Giere, Morrison, and others. One unifying theme in the different terminology of these authors is that while theories are related with general propositions (often in the form of laws of nature), models are less general, but closer to the phenomena (or data). For the purpose of this paper, this vague intuitive understanding of models and theories suffices.

Let us first consider again our simple model of the earth's orbital motion around the sun. This model is derived from Newton's theory of gravitation and Newton's second law. Deriving the model invokes neglecting the internal structure of the two bodies, neglecting other planets or comets, and approximating the motion as a two-body problem with only point masses. As argued by Bailer-Jones (2003), the adequacy of the assumptions made in deriving the model of the earth's motion around the sun is not captured in their propositional content alone. The assumptions are simply wrong; transcribing the "message" of the model solely into propositions would yield to a false model. This, however, does not do justice to scientific practice. Nonetheless, the statement made above reaches further than the observation that the content of a model cannot be equated with its propositions; rather it was claimed that in deriving models it is less knowledge, but abilities that are involved. As stated in section [● Predicting Complex Systems](#) the system of the earth's movement around the sun as considered here is not complex. The message of the model may not be translatable into propositions, but the derivation seems not to involve much skill either. This is different for complex systems.

Note that in the past it may not have been the case that the earth–sun motion is a simple system in the epistemic sense adopted here. The perturbation theory showing the adequacy of the assumptions made, that is, showing that the two-body problem gives a first approximation to the real motion of earth and sun and all other factors may be treated in a perturbation expansion, was only shown more recently. The classification of a system as complex or simple therefore depends crucially on our background knowledge and thus, as indicated in section [● Climate Science or Climate Fiction?](#), may change over time.

As outlined in section [● Climate Science or Climate Fiction?](#), important feedbacks are neglected between various components of the climate system when deriving climate models from more fundamental theories like atmospheric chemistry or hydrodynamics. Our present understanding of the climate system does not allow us to "prove" the correctness or adequacy of all assumptions. Making the right assumptions is, in parts, an ability, a skill that has to be trained rather than a knowledge of facts (broadly understood) that can be learned from text books. Note that this feature of climate modeling is associated with the complexity (in Shackel's sense) of the modeled system and thus holds also for models of other complex systems.

The argument that not only knowledge, but also skill is involved in deriving models representing complex target systems seems easy to buy; however, it has wide-ranging consequences. The assumptions involved in deriving a model may not be correct, but as they are propositions, one may assess the deviations of what is currently perceived as true. For example, one may assess the wrongness of the assumption that both Earth and Sun are point masses

considering the distance between the two bodies. When abilities are centrally involved in deriving a model and these abilities cannot be reduced to propositions, this seems out of reach. Not only can the abilities be interpreted as part of an intricate web of background knowledge; rather they involve knowledge as to how to pursue the specific experimental paradigm, the accepted practice, and the general research experience within the field. These factors cannot be defined explicitly, but must be learned by working in the field. In this respect, a scientific community can be seen as an instance of a Wittgensteinian language community (Wittgenstein 2001, p. 10):

- ▶ [T]he term “language game” is meant to bring into prominence the fact that the speaking of a language is part of an activity [...].

As an example of a language game, Wittgenstein himself refers to “presenting the results of an experiment in tables and diagrams” (Wittgenstein 2001, p. 10). Likewise, assessing the reliability of climate models is, at least to some extent, something that is learnt by the practice of carrying out and verifying such predictions.

As there are borderline cases which are neither clearly complex nor simple, it may remain unclear whether the reliability of some models may be quantified. Nonetheless, there are clear cases for which they can or cannot be quantified; regarding the impacts of the greenhouse effect on humans, the climate system is an example of a complex system and thus the reliability of the models cannot in full be communicated outside the scientific community.

## Reducing Uncertainty

---

As argued above, due to model conceptualization uncertainty, the reliability of complex models cannot entirely be communicated to outside the scientific community. Estimating as to how well a model represents the relevant causal mechanisms if these are still unknown has not much prospect of success. Quantitative figures aiming to capture this reliability are misleading and miss central points. Hence a model pluralistic approach as pursued by the IPCC is also unable to capture all uncertainties because it cannot reflect different reliability of different models.

However, there are other ways of assessing the reliability and thus the uncertainty of forecasts that do not take the circuit via the model's reliability. One may, for example, fall back on a Bayesian account: Via subjective probabilities, the reliability of scientific outcomes may be quantified directly. But it is first impossible to choose a meaningful prior probability due to the large timescales on which the climate system reacts to changes. Second, there is insufficient data for updating these probabilities. The Bayesian method thus fails for climate change. Note that when scientists refer to Bayesian approaches, they quite often estimate a priori probabilities via frequencies and thus by conditional probabilities (conditioned on the models used to determine the respective frequency). These types of Bayesian approaches are, however, model dependent. For further criticisms of the Bayesian approaches, see, for example, Sober (2005).

Another way as to how to assign subjective probabilities may follow Laplace's principle of insufficient reason: All possible effects are taken as equally probable. This approach was put forward, for example, by Harsanyi (1975, 1982). But there is no logical superiority of Harsanyi's assumption of equiprobability over Rawls' focus on the worst outcome as per se there is no logical need to assign subjective probabilities to uncertain decision outcomes based on

Harsanyi's equiprobability assumption or other. We do have information on the likelihood of certain effects of climate change – albeit these are hard to communicate outside one's own narrow scientific community.

Though this chapter is on epistemic issues and thus I do not want to dwell on the interpretation of probabilities in detail here, note that a weak objective interpretation of probability is adopted here: Our most successful method for tackling uncertainty has been to regiment situations of uncertainty by the use of probabilistic propositions. But unless one is a certain kind of subjectivist about probability, one wishes that one's probabilistic beliefs be constrained by objective facts so that they approximate objective variables. The term probability in its everyday use captures something real in the sense that the probability that it rains tomorrow is somehow linked to the real world, in this case the likelihood that it will rain tomorrow. Note that even a modest subjective interpretation of probabilities, which takes probabilities to be belief functions that obey certain restraints, may be consistent with this weak objective interpretation.

As argued above and elsewhere (cp. Frame et al. 2007), there is no reliable basis for assigning probabilities in the above sense to the empirical input needed for a practical assessment. However, there is good practice of how to deal with uncertainties. When the conditions of Bayesianism are not met as in the case of climate change, dealing with uncertainties always invokes the reliability of the model's representation of the considered target system. There is no distinct science as to how to deal with uncertainty; good practice of dealing with uncertainty is always sensitive to the context. It must be learnt by doing and unlike propositional knowledge, it cannot be simply acquired, but the competence must be trained to reach a higher grade. It is beyond the scope of this chapter to address these rather technical issues; the reader is referred to Moss and Schneider (2000), van der Sluijs et al. (2005), and references therein.

## Limits of Common Decision Approaches When Applied to Climate Change

### Interdependencies Between Moral and Epistemic Issues

Summarizing the epistemic part of this chapter a terminology of technology assessment, how to react to global warming is a genuine decision under uncertainty as we lack probability estimates on (at least some of) the possible outcomes. This is distinguished from both decisions under risk in which we do have some reliable probability estimates and from decisions under ignorance where we lack even information on the nature of the possible outcomes. It was argued that presupposing a weak realistic understanding of probability, it is not possible to reduce the global warming issue to a risk problem. This was linked to the fact that predicting the exact impacts of greenhouse gas emissions involves modeling of complex systems. In deriving such models, next to propositional knowledge, certain abilities are of importance. The latter can only be learned by doing, by working in the field. Assessing the “reliability” of certain steps in deriving complex models cannot entirely be communicated to outside the scientific community. A distinction between conceptualization and parameter uncertainty was introduced that may provide the ground for practical decision making to incorporate the uncertainties: While the former is related to the non-propositional content of

a model and thus may not be satisfactorily reflected in a quantitative figure, for the latter quantitative estimates may be given.

All these uncertainties of climate predictions are discussed intensively within the scientific community – not only among climate skeptics. However, uncertainties are often kept under wrap when scientific findings are communicated to the public. Just compare, for example, the full IPCC report and its summary for policy makers (Solomon et al. 2007). It is not the scientists who are to blame here. Rather the practical debate seems incapable of adequately reflecting uncertainties in modeling predictions. If these uncertainties were communicated, sound scientific research runs the risk of being discredited as unscientific; the public seems to prefer black and white instead of the scientists' shades of grey: Often predictions are taken either as correct and unquestionably reliable or simply as wrong. However, most scientific models are neither true, in the sense that they exactly predict future events, nor simply wrong and useless (cp. Giere 2004). The preceding sections aimed to show that uncertainties that cannot be reduced to quantitative figures are necessarily involved in climate modeling. It is argued in this section that in order to incorporate aspects of inter- and intragenerational justice, practical decision making has to carefully consider the shades of grey that affect the reliability of climate models in practical decision making. Thereby, the focus of this section is not, like in Gardiner (2006a), Hanson and Johannesson (1997), Lumer (2002), and others, on the question of how much, if any, reduction of greenhouse gases is ethically legitimate, but rather on what kind of decision-making criteria should guide our reasoning about this very question. Thereby, a very simplified and rather unrealistic setting is considered, namely, that a well-defined decision maker (or a group of well-defined decision makers) dealing with climate change issues exists.

Why, after all, do we have to worry about epistemic problems when reasoning about issues of inter- and intragenerational justice? You may wonder why we cannot – even if unquantified uncertainties are a problem in principle and cannot be evaded even with future research – just simply wait until climate models and global and long-term economic predictions have overcome their teething troubles. For sure, climatology as well as impact modeling are huge scientific research fields that have advanced a great deal over the last years – and most likely will do so in the near future. However, practical decision making facing the threats of climate change cannot wait for better prognoses. The climate system only reacts very slowly to changes in its parameters, such as changes in carbon dioxide contraction. Hence the atmospheric concentration of persistent greenhouse gases like carbon dioxide can only be stabilized by reducing emissions (Solomon et al. 2007). The large inertia of the climate system necessitates timely countermeasures. Once particular effects occur, it may well be too late for a systematic response. Note again that this chapter deals with a sound discourse about how to react to climate change, not with the issue of correct reduction, mitigation or adaption strategies. The practical discourse may or may not come to the conclusion that instead of mitigating now, we should wait and adapt later. However, this decision cannot wait for better and less uncertain predictions: it has to be taken now.

A need to address epistemic uncertainties in practical debates can be deduced from three (fairly weak) assumptions: First, practical decision making has to be based on the best (empirical) knowledge available. Second, practical problems related to environmental issues can be formulated as scientific problems. Third, science gives us the most reliable understanding of the natural world. There seems to be no need to justify these suppositions, as all three

seem to be both weak and rather intuitive. From these epistemic and practical assumptions, it follows that we have to consider epistemic uncertainties in practical decision making: The best available information that we have today is our scientific forecasts *plus* information on their reliability. Though the latter may not be expressed or even be able to be expressed in numeric terms, information on the quality of various climate predictions is available. If, for example, quantified uncertainties that arise from insufficient knowledge on the input parameters were the only uncertainties we had to deal with, common probabilistic decision criteria like utility maximization could be applied in a straightforward way. As argued above, unquantified uncertainties, however, that arise from insufficient understanding of the model conceptualization pose a severe problem. Quantitative figures may be misleading, but they can be communicated easily to people outside one's own discipline. Even if scientists in a given field tend to assign "higher-order beliefs" that express their confidence in an underlying theory, the methodology used, the researcher or the group who carried out the work, etc., these higher-order beliefs are only very rarely quantifiable themselves in terms of, say, subjective probabilities.

## Expected Utility Maximization and the Precautionary Principle

---

The lack of probability estimates on the level of impact models renders a maximized utility approach impossible. Maximizing the expected utility is an adaption of the utilitarian maxim of the *greatest good for the greatest number* to decisions under uncertainty: It is not the overall utility (or "good") that is to be maximized, but the expected utility, that is, the sum of different utilities weighted by their probability of occurrence. The assignment of utilities to possible climate-change effects raises many difficult problems, but I do not want to dwell on them here. These problems are not specific to decision making under uncertainty or even related to problems of welfare-based ethics. In particular, the problems about determining the utility of an event, or deciding what utility actually amounts to, parallel to some extent problems of the precautionary approach discussed below when deciding as to how to actually determine the worst-case scenario. Note that nonetheless the problem of the precautionary approach is somehow easier, as it needs only an ordinal concept of well-being, while an expected utility approach presupposes a cardinal welfare measure. Only when the welfare function fulfills certain restraints, ordinal measures may be mapped onto cardinal ones (von Neumann and Morgenstern 1947).

As this paper's focus is on uncertainties (of expected utility and of the worst-case scenario), the problems associated with measuring human welfare and how to equate it with utilities are not discussed here. For the purposes of this paper, it suffices to assume that the impact on human welfare estimated in economic models on level 3 in  Fig. 13.1 can be associated with (intersubjective) utilities in a meaningful way. How to actually assign meaningful utilities has been discussed extensively in the literature. It raises rampant problems particularly for intergenerational ethics; see, for example, Lumer (2002) for a discussion as to how assign utilities in the context of climate change.

When outcomes are highly uncertain, it is often suggested that we fall back on the precautionary principle. This principle was first made popular within environmental ethics by H. Jonas in his *imperative of responsibility* in the late 1970s. As the phrase *precautionary*

*principle* is fraught with ambiguity, let us briefly explicate the term and its use within ethical, juridical, and political contexts. The Rio Declaration on Environment and Development, for example, formulates the precautionary principle (rather vaguely) as follows (cp. UNFCCC 1992):

- ▶ Where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation.

In this weak formulation, the precautionary principle provides no distinct directive for practical decision making. Instead it constitutes a meta-criterion stating that uncertainties in scientific forecasts have to be taken seriously. Strong formulations of the precautionary principle constitute a genuine decision criterion. The following is an example of the *precautionary principle* in a strong formulation (The Wingspread Statement 1998):

- ▶ Where an activity raises threats of harm to the environment or human health, precautionary measures *should be taken* if some cause and effect relationships are not fully established scientifically [my italics].

Note that apart from the two versions discussed in this paper, various other formulations of the precautionary principle exist (cp. Sandin et al. 2002). Proponents of the precautionary principle like C. Raffensberger and J. Tickner suggest the following core idea behind all formulations of the *precautionary principle* (Raffensberger and Tickner 1999, p. 1):

- ▶ In its simplest formulation, the precautionary principle has a dual trigger: If there is a potential for harm from an activity and if there is uncertainty about the magnitude of impacts or causality, then anticipatory action should be taken to avoid harm.

In this chapter, the precautionary principle is understood as a genuine decision-making criterion that, loosely following Gardiner (2006b), interprets the strong formulation as a variant of the minimax rule in decision theory: Minimize the maximally bad outcome. Given certain assumptions about how to quantify harm and well-being, this may be reformulated as a maximin rule and reads (for climate change): Maximize well-being in those scenarios in which the involved humans are worst off (minimally benefited), regardless of how uncertain these scenarios are.

At first glance, a precautionary approach seems well suited to avoiding an ethically unjustifiable discounting of future damage caused by our present greenhouse gas emissions: We cannot exclude with certainty the possibility that the release of greenhouse gases has the potential to cause severe harm to future generations; hence emissions of greenhouse gases ought to be abandoned. A precautionary approach seems adequate when the stakes are high – the living conditions of all future humans may be endangered by severe climatic changes. Though there is considerable disagreement within the economic community on the costs of reducing greenhouse gases (cp. the response of Weitzman (2009) and Nordhaus (2008) to Stern (2007)) there are some economic assessments suggesting that reducing anthropogenic greenhouse gas emissions is not very costly. Following Stern (2007), a commitment of only 1% of global gross domestic product (GDP) is needed to avoid the major hazards that may arise from climate change. At first blush, this appears very affordable; but if we base our calculation on current GDP value, it amounts to an investment of US\$450 billion per year. For comparison:

the current estimates of the money needed to provide 80% of rural populations in Africa with access to water and sanitation by 2015 amounts to US\$1.3 billion per annum (cp. Martinez Austria and van Hofwegen 2006). Clearly, societies (or other organizations) are able to part with only a certain amount of money or other resources for altruistic endeavors, and the mitigation of major changes in future climate is only one such endeavor. Investing in the mitigation of climate-change effects means forgoing other investments which we have a moral obligation to make. One central requirement of the practical debate is a decision about which investment has priority over others. Presupposing an answer to this question from the very beginning of the debate – for example, by assuming that climate change is currently humanity's most pressing problem – preempts the moral debate, cp. [Fig. 13.1](#). Applying the precautionary principle to global warming as a singular problem thus does not allow us to adequately deal with the valid claims of groups that are adversely affected by natural or societal “disasters” other than climate change. This approach is clearly incapable of adequately incorporating considerations of inter- or intragenerational justice as it does not address the question as to why suffering arising from climate change has priority over suffering caused by other sources.

The critique raised here, however, is not a charge against the precautionary principle itself; it only disqualifies common applications of the precautionary principle. Suppose that one can show that – given certain ethical standards, which are not under debate here – the worst-case scenario regarding the effects of climate change is that these effects are worse than any other type of human suffering, present and future. This means that if our (possibly very unrealistic) assumptions are indeed correct, then following the precautionary principle, we have to mitigate climate change at any cost. We are trading the certain suffering of people living presently against a possibly more severe, but yet uncertain suffering of people living in the future. If the worst-case scenario is as uncertain as currently estimated for global warming, and is balanced against certain other scenarios whose bad effects are certain (like the actual suffering of many people in third world countries, for example), it is unreasonable to completely mask all other scenarios and focus on mitigation of the uncertain, but worst outcome. Note that this argumentation needs refinement when the worst-case outcome is a singular event like the end of human life on Earth (cp. Ord et al. 2009). Current empirical evidence seems to exclude the possibility that climate change is of this severity.

As noted above, the available information on the effects of anthropogenic global warming includes information about the “likelihood” of the worst-case and other scenarios. This information is not quantified and may not be fully quantifiable at all. However, we do have information that suggests that, while present suffering is certain, future suffering caused by global warming is uncertain. Good arguments for neglecting this information should be given. But to the best of the author's knowledge, no such arguments have been presented in the literature. Problems with a precautionary approach as an action-guiding principle have been discussed extensively in the literature (cp. Peterson 2006; Clarke 2005, and references therein). This paper only addresses one central issue of importance in any intergenerational ethics, namely, how to balance obligations toward future generation against obligations toward people living presently. Even if one argues for the ethical legitimacy of trade-offs between losses and gains experienced by different people, one cannot deny that presently living people have a right to safe water and sufficient nutrition. One needs to argue at least that uncertain future losses are worse than the current suffering. But a mere precautionary

approach to global warming is incapable of simultaneously incorporating considerations of inter- and intragenerational justice.

## Summary and Further Research

---

Sections [Climate Science or Climate Fiction?](#), [Predicting Complex Systems](#), [Knowing-How in Scientific Modeling](#) focused on epistemic aspects of climate and impact models and brought forward the argument that not all uncertainties involved in modeling complex systems may be adequately captured by probabilities. Arguing that not all aspects of scientific uncertainties may be captured in numerical figures and that they may be misunderstood outside the scientific community is not to be mistaken as putting the case for a dictatorship of experts. Rather it puts pressure on the scientists to provide standards of good practice. If anything, this chapter aims to invoke not to trust any numerical figure only because its predictions and uncertainties are captured in quantitative terms – its prognoses may not be better or more scientific than qualitative ones. In addressing the uncertainty of climate predictions from an epistemic view, this paper wishes to provide a basis for incorporating all the scientific findings into the practical debate, that is, climate or impact model predictions plus the associated uncertainty including those that may not be quantified.

As argued in section [Limits of Common Decision Approaches When Applied to Climate Change](#), the precautionary principle, due to its sole focus on the worst-case scenario, is incapable of dealing with issues of inter- and intragenerational issues as they are relevant in the climate debate in an adequate way. Instead of focusing on the worst-case scenario, an expected utility approach considers all possible outcomes and the associated utilities, weighted by their occurrence probability, or, to put it in more technical terms, we are to maximize and take into account all possible scenarios. The extreme scenarios of runaway climate change or very little temperature change, for example, are thus taken into account, as is the scenario in which the temperature change exactly equals the estimated mean value. The latter scenario being the most probable is given the greatest weight. However, an expected utility approach is not applicable for climate change in a straightforward way as there are no reliable probability estimates at hand on the level of impact modeling. This is unfortunate as maximizing expected utility has one clear advantage over a precautionary approach: By incorporating an inter-temporal as well as an international perspective, maximizing expected utility is, by its very nature, able to trade-off the costs and benefits of different people living at different places and times.

The lack of (subjective) probabilities in the objective sense defined above does not imply, however, that one has to fall back on non-probabilistic decision criteria such as the precautionary approach. This chapter's proposition should not be misunderstood as a kind of reformulation of the precautionary principle in its weak form, that is, "Take uncertainties seriously and therefore address also the uncertain outcomes." Rather this contribution aims to argue that uncertain effects are not to be (mis)taken as certain ones, which seriously undermines the use of the precautionary principle.

In the literature, decision methods are suggested, which parallel expected utility maximization, to cope with the lack of reliable prior probabilities and information about how to update these priors on the basis of the conditional probability calculus (e.g., Shafer 1990).

An adequate decision procedure for global warming would assign meaningful utilities to various outcomes in a first step by political decision makers, moral philosophers, and others. As to the occurrence of unquantified uncertainties, however, the second step, the actual cost-benefit analysis (understood in a broad sense), should be conducted by experts on the empirical forecasts. Without expecting a detailed casuistic from the philosophical analysis, such a blueprint can only work when philosophical ethics does not shy away from context-variant information on the very decision. The 1970s debate on the “rationality” of expected utility maximization or maximin, whose main protagonists were Harsanyi and Rawls, shows that answering the question of whether the precautionary principle or expected utility maximization is adequate has to, willy-nilly, implement context-variant features of the decision situation. Note that this paper argues against the precautionary principle only when applied to global warming. The given arguments do not discredit this principle as a decision-making criterion in itself. Concerning an adequate decision-making approach to global warming, this chapter has, so far, turned a blind eye to factors that actually precede the debate on whether the precautionary approach or expected utility measures seem most adequate. So, concluding this paper, let me briefly touch on this problem: Before being able to actually talk about uncertain outcomes of the decision to reduce (or not reduce) greenhouse gases, we have to decide what this decision is actually all about – is it about the welfare of future humans?, do we need to discuss the pros and cons of alternative energy supplies that do not emit greenhouse gases as well?, etc. Any analysis of a specific decision must start with some delimitation of the decision itself. It is not always well established how to determine the “decision horizon” (Hansson 1996, p. 371): The scope of the decision, or even which problem the decision is supposed to solve might be unclear. The further in time the consequences of our decisions lie, the more difficult it is to determine the decision horizon.

Currently, the decision horizon is most often set by pragmatic considerations only, however there is more systematic work for philosophers to be done here. It seems that here – just like for the question as to what decision criteria can adequately deal with uncertainties – rule-based approaches are at their limits, both at the ethical as well the epistemic level. Investigating of how virtue ethics (cp. Luntley 2003) or virtue epistemology, particularly virtue responsibilism, may help to solve these problems seems thus a worthwhile endeavour. Virtue approaches by their very nature are capable of incorporating the vagueness and uncertainties of decision situations by fostering certain dianoetic, that is, epistemic virtues. An example is given by the Aristotelian phronesis that provides the moral agent with the mental capacity to adequately judge the unquantified uncertainties connected to the scientific predictions. Virtue epistemology may provide new insights into the field of risk research that supplement the debate on how moral virtues and moral emotions can help in an adequate response to risks and uncertainties such as those posed by global warming.

## Acknowledgments

---

I am particularly grateful to my colleagues Steve Clarke, Peter Taylor, Nick Shackel, and Martin Peterson for constructive input on the issues discussed in this chapter and comments on (earlier versions of) this manuscript. I thank Kai Hennes for help in finalizing the manuscript. Parts of this chapter were published in Hillerbrand (2010, 2009).

## References

---

- Bailer-Jones DM (2003) When scientific models represent. *Int Stud Philos Sci* 17(1):59–74
- Betz G (2009a) Underdetermination, model-ensembles and surprises: on the epistemology of scenario-analysis in climatology. *J Gen Philos Sci* 40(1): 3–21
- Betz G (2009b) What range of future scenarios should climate policy be based on? Modal falsificationism and its limitations. *Philos Nat* 46(1):133–158
- Clarke S (2005) Future technologies, dystopic futures and the precautionary principle. *Ethics Inf Technol* 7:121–126
- Devall W, Sessions G (1985) Deep ecology: living as if nature mattered. G. M Smith, Salt Lake City
- Frame DJ, Faull NE, Joshi MM, Allen MR (2007) Probabilistic climate forecasts and inductive problems. *Philos Trans R Soc A*. doi:10.1098/rsta.2007.2069
- Gardiner SM (2006a) A core precautionary principle. *J Polit Philos* 14(1):33–60
- Gardiner SM (2006b) A perfect moral storm: climate change, intergenerational ethics and the problem of moral corruption. *Environ Values* 15:397–413
- Giere R (2004) How models are used to represent reality. *Philos Sci* 71:742–752
- Hansson SO, Johannesson M (1997) Decision-theoretic approaches to global climate change. In: Fermann G (ed) International politics of climate change. Scandinavian University Press, Stockholm, pp 153–178
- Hansson SO (1996) Decision-making under great uncertainty. *Philos Soc Sci* 26:369–386
- Harsanyi JC (1975) Can the maximin principle serve as a basis for morality? A critique of John Rawls' theory. Reprinted in: Richardson HS, Weithman PJ (eds) The philosophy of John Rawls. A collection of essays. Taylor & Francis, Cambridge, pp 234–246
- Harsanyi JC (1982) Morality and the theory of rational behaviour. In: Sen A, Williams B (eds) Utilitarianism and beyond. Cambridge University Press, Cambridge
- Hillerbrand R (2009) Epistemic uncertainties in climate predictions. A challenge for practical decision making. *Intergenerational Justice Rev* 3:94–99
- Hillerbrand R (2010) On non-propositional aspects in modelling complex systems. *Analyse & Kritik* 32:107–120
- Hillerbrand R, Ghil M (2008) Anthropogenic climate change: scientific uncertainties and moral dilemmas. *Physica D* 237:2132–2138
- Houghton RA (1995) Effects of land-use change, surface temperature and CO<sub>2</sub> concentration on terrestrial stores of carbon. In: Woodwell GM, Mackenzie FT (eds) Biotic feedbacks in the global climatic system: will warming feed the warming? Oxford University Press, New York, pp 330–350
- Idso SB (2001) Carbon-dioxide-induced global warming: a sceptic's view of potential climate change. In: Gerhard LC, Harrison WE, Hanson BM (eds) Geological perspectives of global climate change. American Association of Petroleum Geologists, Tulsa, pp 317–336
- IPCC (2007) Summary for policymakers. In: Solomon et al. 2007
- Krebs A (1999) Ethics of nature. A map. de Gruyter, Berlin/New York
- Kuhlmann M (2010) Mechanisms in dynamically complex systems. In: McKay Illari P, Russo F, Williamson J (eds) Causality in the sciences. Oxford University Press, Oxford
- Lumer C (2002) The greenhouse. A welfare assessment and some morals. University Press of America, Lanham/New York/Oxford
- Luntley M (2003) Ethics in the face of uncertainty: judgement not rules. *Bus Ethics Eur Rev* 12(4):325–333
- Mackenzie FT, Lerman A, Ver LM (2001) Recent past and future of the global carbon cycle. In: Gerhard L, Harrison WE, Hanson BM (eds) Geological perspectives of global climate change. American Association of Petroleum Geologists, Tulsa, pp 51–82
- Martinez Austria P, van Hofwegen P (eds) (2006) Synthesis of the 4th World Water Forum. National Water Commission of Mexico, Mexico City
- Moss RH, Schneider SH (2000) Uncertainties in the IPCC TAR: recommendations to lead authors for more consistent assessment and reporting. In: Pachauri R, Taniguchi T, Tanaka K (eds) Guidance papers on the cross cutting issues of the third assessment report of the IPCC. World Meteorol. Org., Geneva, pp 33–51
- Nordhaus W (2008) A question of balance. Weighing the options on global warming policies. Yale University Press, Yale
- Nordhaus WD, Boyer JG (2000) Warming the world: the economics of the greenhouse effect. MIT Press, Cambridge, MA
- Ord T, Hillerbrand R, Sandberg A (2009) Probing the improbable: methodological challenges for risks with low probabilities and high stakes. *J Risk Res* 13(2):191–205
- Oreskes N (2004) Science and public policy. What's proof got to do with it? *Environ Sci Policy* 7(5):369–383

- Parker WS (2006) Understanding pluralism in climate modeling. *Found Sci* 11:349–368
- Peterson M (2006) The precautionary principle is incoherent. *Risk Anal* 26(3):595–601
- Polanyi M (1967) The tacit dimension. Anchor Books, New York
- Raffensberger C, Tickner J (eds) (1999) Protecting public health and the environment: implicating the precautionary principle. Island Press, Washington, DC
- Refsgaard JC, van der Sluijs JP, Brown JD, van der Keur P (2006) A framework for dealing with uncertainty due to model structure error. *Adv Water Resour* 29:1586–1597
- Sandin P, Peterson M, Hansson SO, Rudén C, Juthe A (2002) Five charges against the precautionary principle. *J Risk Res* 5:287–299
- Shafer G (1990) Perspectives on the theory and practice of belief functions. *Int J Approx Reason* 4:323–362
- Sober E (2005) The design argument. In: Man WE (ed) *The Blackwell guide to the philosophy of religion*. Blackwell, Malden/Oxford/Carlton
- Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) (2007) The physical science basis. Contribution of WG I to the 4th assessment report of the IPCC. Cambridge University Press, Cambridge/New York
- Stern N (2007) The economics of climate change: the stern review. Cambridge University Press, Cambridge
- Thoreau HD (2004) *Walden and other writings*. Elibron Classics, Chestnut Hill
- Tol RSJ (2002) Estimates of the damage costs of climate change. Part II: dynamic estimates. *Environ Resour Econ* 21:135–160
- United Nations Framework Convention on Climate Change (UNFCCC) (1992) Kyoto protocol to the United Nations Framework on Climate Change. <http://unfccc.int/resource/docs/convkp/kpeng.pdf>
- van der Sluijs JP (1997) Anchoring amid uncertainty; on the management of uncertainty in risk assessment of atherogenic climate change. Ph.D. thesis, Utrecht University
- van der Sluijs JP, Risbey JS, Kloprogge P, Ravetz J, Funtowicz SO, Corral Quintana S et al. (2003) RIVM/MNP guidance for uncertainty assessment and communication: detailed guidance, report commissioned by RIVM/MNP – Copernicus Institute, Defrayment of Science, Technology and Society. Utrecht University, Utrecht
- van der Sluijs JP, Craye M, Funtowicz SO, Kloprogge P, Ravetz J, Risbey J (2005) Combining quantitative and qualitative measures of uncertainty in model based environmental assessment: the NUSAP system. *Risk Anal* 25(2):481–492
- von Neumann J, Morgenstern O (1947) *Theory of games and economic behavior*, 2nd edn. Princeton University Press, Princeton
- Walker WE, Harremoes P, Rotmans J, Sluijs JP, van Asselt MBA, Janssen PH, Krayer von Krauss MP (2003) Defining uncertainty. A conceptual basis for uncertainty management in model-based decision support. *Integr Assess* 4:5–17
- Weitzman ML (2009) On modeling and interpreting the economics of catastrophic climate change. *Rev Econ Stat* XCI(1):1–19
- Wilson RC, Drury SA, Chapman JL (2000) *The great ice age*. Routledge, London/New York
- Wittgenstein L (2001) *Philosophical investigations*. Blackwell, Oxford
- Yancheva G, Nowaczyk NR, Mingram J, Dulski P, Schettler G, Negendank JFW, Liu J, Sigman DM, Peterson LC, Haug FH (2007) Influence of the inter-tropical convergence zone on the East Asian monsoon. *Nature* 445:74–77



# 14 Earthquakes and Volcanoes: Risk from Geophysical Hazards

Amy Donovan

University of Cambridge, Cambridge, UK

<b>Introduction: Lithospheric Risk .....</b>	<b>342</b>
Defining Hazard, Risk, and Uncertainty in Disaster Response .....	343
Geography of Risks and the Problem of Induction .....	345
Summary: Categorizing Uncertainty .....	348
<b>History .....</b>	<b>350</b>
Earthquake Risk .....	351
Volcanic Risk .....	355
Summary .....	357
<b>Current Research .....</b>	<b>357</b>
Scientific Uncertainty .....	358
Earthquakes .....	358
Volcanoes .....	360
Social Aspects of Risk and Uncertainty .....	361
Risk Perception and Communication .....	361
Risk Governance and Policy Making .....	364
Summary .....	364
<b>Further Research .....</b>	<b>365</b>
Uncertain Science .....	365
Uncertain Societies .....	366
<b>Conclusions .....</b>	<b>367</b>

**Abstract:** The aims of this chapter are to present a brief history of ideas in the interdisciplinary study of volcanic and seismic risk, to discuss the current state of the subject, and to suggest pathways for further research. This is a very extensive topic – while much of the scientific literature tends to focus on hazard assessment (and, increasingly, risk assessment), the social sciences have tended to focus on vulnerability reduction and risk communication. There have been very few holistic epistemological studies of the broader context of risk. Yet the philosophical aspects of uncertainty are increasingly important for scientists in particular as they seek to assess and understand these risks, not least because of heated debates within both fields concerning the relative values of deterministic and probabilistic methods and the ways in which they deal with uncertainty. Social scientific and philosophical methods therefore have significant potential to inform this discussion, and are also increasingly important in assessing vulnerability and popular understanding of risks in hazardous areas. There has been a large volume of work done in recent years to examine seismic and volcanic risk perception and communication, much of which suggests that these risks are not high on the social agenda until an event happens. This calls for new approaches to population management, preparedness, and proactive roles for scientists and social scientists.

## Introduction: Lithospheric Risk

---

In recent years, population growth around the world has led to huge increases in the number of people at risk from geophysical hazards (taken in this chapter to refer to those associated with earthquakes and volcanoes – perhaps more accurately risks from within the lithosphere). It is estimated that over half a million people have died as a result of earthquakes in the last 10 years (Spence 2009), and that almost a billion are at risk from volcanic activity (Ewert and Newhall 2004). Globalization, raised life expectancy, and indeed technological development have all contributed to the increased risk: While the “risk society” (Beck 1992, 1999, 2009) has largely been associated with technological risk, the amplification of natural risks is an unfortunate side effect – and one that is often overlooked in the risk theory literature. Yet natural risks have been increased as a result of population growth – which can be linked to medical advancements and technological enhancements to the quality of life. There are also complications arising from technologies like aviation, demonstrated dramatically by the 2010 “ash crisis”: The eruption in Iceland was relatively small, but the vulnerability of aircraft significantly increased its impact (Donovan and Oppenheimer 2010).

Beck himself, in *World at Risk*, wrote that “even though human interventions may not be able to prevent earthquakes or volcanic eruptions, these can be predicted with reasonable accuracy” (p. 50): a widespread opinion among the general population in the West, but one that is not backed up by the state of the science. Indeed, some seismologists do not believe that it is possible to predict earthquakes (e.g., Geller 1997; Matthews 1997), and volcanologists cannot predict eruptions (Sparks 2003). Scientific understanding of volcanic and seismic processes is advancing rapidly, but the nature of the events is such that even with greater knowledge, the uncertainty about the natural system is very high. Monitoring events deep in the earth is extremely challenging and often logically impossible – scientific methods are largely, at present, dependent on measurements made in the upper crust.

The volcanic island of Montserrat in the West Indies awoke on July 18, 1995, to explosive activity from its long-silent volcano (Druitt and Kokelaar 2002). Over the following few

months and then years, Montserratians learned that science could not give them the answers that they needed – their capital city was evacuated for the third and final time in 1996, and now sits under several meters of pyroclastic debris. A key breakthrough in the communication of risk on Montserrat was simply that the science was very uncertain: Relationships between scientists, between scientists and the public, and between scientists and policymakers were particularly tense in the early years of the eruption in part as a result of high expectations of science (Pattullo 2000; Donovan 2010). The scientific community responded by developing new methods for risk assessment. The Montserratian community responded either by learning to live with the volcano, or by relocating. Both communities (and the many subgroups within them) have been changed by the events.

## Defining Hazard, Risk, and Uncertainty in Disaster Response

In natural hazards research, variations of the conceptual formula,  $\text{risk} = \text{hazard} \times \text{vulnerability}$  are frequently used to distinguish between the natural event and its impact on populations (e.g., Wisner et al. 2004; Wang 2009). This has frequently led to a decoupling of research cultures, with physical scientists working on hazard assessment and modeling, and social scientists working on vulnerability mapping, risk perception and communication studies, and resilience-building. Engineering studies in seismic hazard management have bridged this gap to some extent, seeking to develop mitigation measures of various kinds, such as earthquake-resistant housing. There is then a further division, between those who do the research and those to whom it might be of use – and this division may be characterized by widely differing perceptions of the importance of the topic, since low-probability, high-impact hazards rarely impact daily life unless the event occurs (Gaillard 2008). This is particularly true in developing countries, where the need to survive on a daily basis precludes longer-term planning. In this case, there is a very important role for NGOs and international organizations such as the UN, which specialize in disaster management. However, communication between these institutions and academic researchers is often poor.

Given the lack of predictive skill in geophysical hazard assessment, much work is focused on responses to events – reducing vulnerability and increasing resilience preemptively can prepare communities to maximize their capacity to survive geophysical hazards. The Hyogo Framework for Action (HFA), adopted in 2005, identified the following areas for concentration (shown graphically in Fig. 14.1):



Fig. 14.1  
Disaster response chain

- Ensure that disaster risk reduction is a national and a local priority with a strong institutional basis for implementation.
- Identify, assess, and monitor disaster risks and enhance early warning.
- Use knowledge, innovation, and education to build a culture of safety and resilience at all levels.
- Reduce the underlying risk factors.
- Strengthen disaster preparedness for effective response at all levels.

Relatively little of this relates to the purely scientific realm of understanding the physics and chemistry of active volcanoes and faultlines. It has rather to do with the ways in which different stakeholders respond to the risks posed by natural phenomena. Governments, scientific assessment, social scientists, NGOs, and local communities are all involved in reducing the risk from natural disasters. This raises social, cultural, philosophical, economic, and epistemological issues that require a transdisciplinary approach that is context-sensitive. Nevertheless, the importance of scientific research into the physical processes involved in generating geophysical hazards, and into modeling them, should not be underemphasized: researchers are still mapping faults and volcanic systems, and uncovering new risks. A further challenge for these researchers is to ensure that their research is appropriately translated for stakeholders in relevant areas: there is a very real danger that research will enter the published literature of the academic community without impacting at local and national levels in government. This was demonstrated on Montserrat: an academic risk assessment of the volcano was carried out in the mid-1980s, but failed to penetrate government agencies, in large part because of its format as a scientific document (Wadge and Isaacs 1988).

Where warning is given of a likely large earthquake, there are several mitigation measures that can be put in place. The most widely recognized is the application of building codes, which typically outline the minimum acceptable precautions that have to be taken in the construction of particular buildings. For example, a hospital might have to be built with steel-reinforced walls, while residential accommodation might be exempt from such costly considerations. Building codes are compulsory in some areas of the world, such as California, where earthquake risk is high. However, there are complexities involved in their enforcement in areas where building techniques are still highly culturally controlled, as in rural Iran, for example. Building codes force contractors to design buildings that are able to absorb some of the shaking from the earthquake, thus making them less likely to collapse.

In volcanic eruptions, engineering solutions are less well developed, partly because there are multiple hazards involved. There have been attempts to redirect lava flows, for example, but these have been erratic in their success. One area that has been relatively productive is the building of channels for mudflow (lahar) redirection: Volcanic mudflows may occur for years after the end of an eruption, when heavy rain remobilizes volcanic ash and debris around the volcano. They tend to be very powerful and very destructive, but may follow predictable paths, allowing for some mitigation measures to be put in place (e.g., Tayag and Punongbayan 1994). Another volcanic hazard where encouraging developments are being made is that of ash and tephra fall. Structural reinforcement of roofs in areas of volcanic risk can lower the risk of roof collapse during ashfall (e.g., Spence et al. 2005). However, this has been done in relatively few locations as yet, and is limited to moderate-sized eruptions. Other volcanic hazards are more problematic, most notably perhaps pyroclastic flows, which are very mobile, fast, and hot, destroying everything in their paths. While some attempts to construct flow-proof buildings

**Table 14.1****Hazards from earthquakes and volcanoes**

Hazards from volcanoes	Hazards from earthquakes
Lava flows	Ground shaking
Pyroclastic flows	Liquefaction
Ash clouds	Building destruction
Ballistics (lava bombs)	Fires
Tsunami	Ground rupture
Earthquakes	Tsunami
Jokulhlaups (flooding due to glacial meltwater)	Floods
Gas and aerosols (e.g., can cause climate forcing)	
Lahars (volcanic mudflows)	
Lightning	
Blasts (e.g., at Mount St Helens in 1980)	

have been made, they have not as yet been adequately tested (e.g., Spence et al. 2008). In summary, **Table 14.1** shows a taxonomy of seismic and volcanic hazards.

## Geography of Risks and the Problem of Induction

Risk from geophysical hazards is strongly geographically constructed, both humanly and physically (e.g., Jackson 2006). Clearly, the risk from volcanoes is amplified in areas where there are volcanoes – although as the 2010 eruption of Eyjafjallajökull demonstrated, volcanic eruptions can affect much larger areas. For very large eruptions, this can extend to global climactic impacts, as noted above. Volcanism and seismicity tend to occur at plate boundaries, where the relative motion of plates causes melting and faulting (**Fig. 14.2**). However, there are important exceptions to this. Intraplate volcanism can occur either at hot spots (mantle plumes) like Hawaii, the Galapagos Islands, and the Azores or at rift zones like the East African Rift. Earthquakes may also occur in these locations, and at other areas like the New Madrid region in the central USA, which is thought to be a failed rift system (e.g., Johnston and Nava 1985; Cox et al. 2001). Additionally, small earthquakes may occur throughout tectonic plates due to stress changes in the crust, and isostatic rebound. Since earthquakes occur around the world very frequently, global seismic hazard based on the distribution of large earthquakes through time can be estimated and mapped (**Fig. 14.3**). However, in seismology there is considerable debate about the relationship between the temporal and spatial distribution of earthquakes, their magnitude, and the physical properties of a particular fault. This discussion will be detailed below.

In volcanology, the problem of forecasting the next eruption is compounded by several complexities, not least the wide variety of types of volcano: Some volcanoes tend to produce evolved, viscous lavas, which often erupt explosively as gas is trapped and becomes pressurized, while others produce relatively benign lava flows. In addition, volcanic activity is the result of

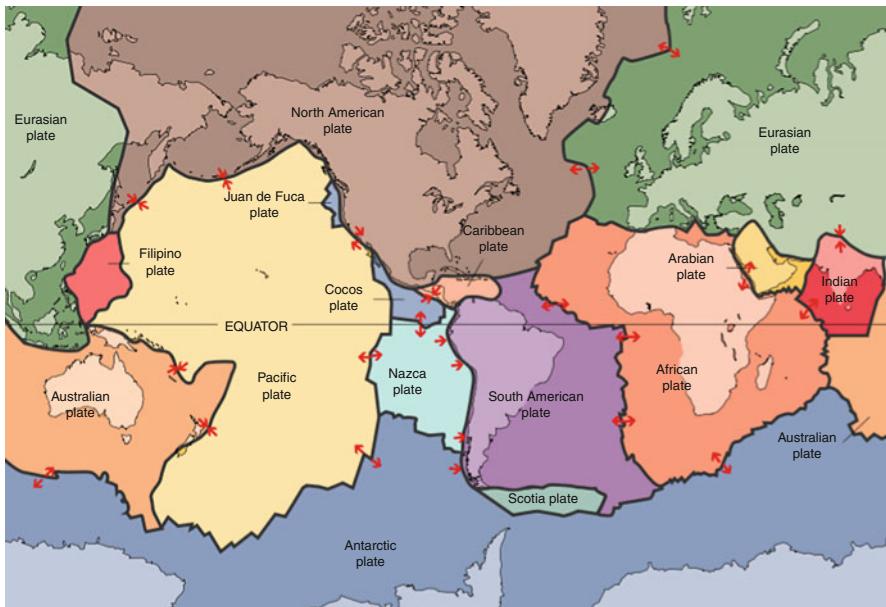


Fig. 14.2

Global tectonic map, courtesy of the US Geological Survey

#### GLOBAL SEISMIC HAZARD MAP

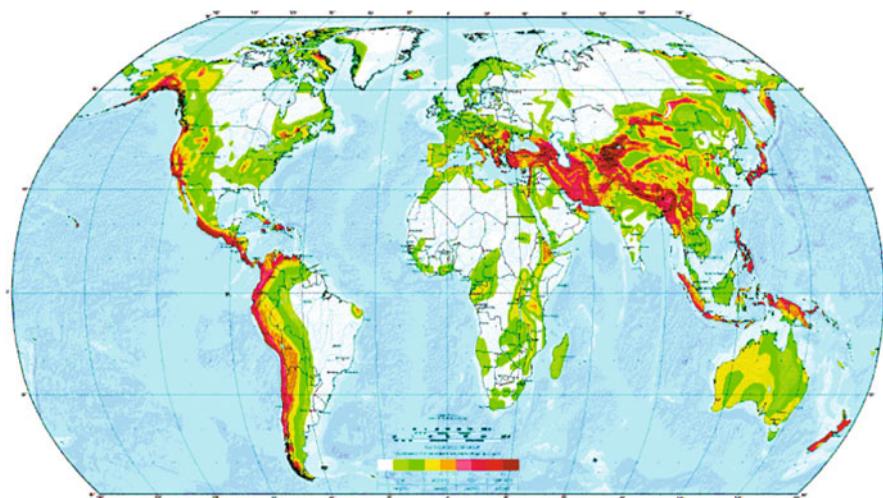


Fig. 14.3

Global seismic hazard map, courtesy of the Global Seismic Hazard Assessment Program. The Global Seismic Hazard Assessment Program (GSHAP) was launched in 1992 by the International Lithosphere Program (ILP) with the support of the International Council of Scientific Unions (ICSU), and endorsed as a demonstration program in the framework of the United Nations International Decade for Natural Disaster Reduction (UN/IDNDR)

a combination of highly nonlinear geophysical and geochemical processes, which may have localized geological controls. At a single volcano, an eruption may alternate between explosive and effusive phases, for example. The fact that relatively few eruptions have been observed with modern technologies renders prediction and forecasting very challenging, particularly given the long periods of repose between eruptions at potentially dangerous volcanoes. Volcanologists are very reluctant to compare volcanoes, at least quantitatively: There is a geographical problem of induction. Extrapolating the behavior of one volcano because of the way a similar volcano has behaved in the past has been shown repeatedly to involve dangerous assumptions – although it is often necessary.

• *Table 14.2* shows some of the key parameters in defining and handling earthquakes and eruptions, and gives some examples of the complexities involved. These can have scientific, social, economic, and cultural implications, and are often highly geographically and indeed historically specific. At the same time, the global nature of societal interactions in the “risk society,” and the structure and resourcing of academic research, means that many of those involved in managing a crisis will be from overseas: on Montserrat, for example, many scientists came from the UK and the USA. In volcanic crises in particular, many volcanologists may wish to be involved, not least because eruptions are relatively rare, spectacular, and present data-gathering opportunities, and the chance to apply one’s expertise for social benefit. Local scientists in developing countries may be dependent upon collaborations with outside scientists for state-of-the-art equipment and labor time (though this can also create some tensions where conflicting agendas are being pursued). This is particularly true on volcanoes, where expensive monitoring networks are simply not available to poorer countries (Donovan 2010).

The impact of global social inequalities on disaster risk has already been touched upon. Developing countries are very vulnerable to natural disasters, especially where populations are expanding rapidly. In Caracas, Venezuela, for example, there is a very high seismic risk because the city is built in an alluvial basin, and this amplifies the shaking from earthquakes. Housing

■ **Table 14.2**

**Key parameters in calculating and managing geophysical hazards and examples**

Problem	Volcanoes	Earthquakes
Duration	Chronic crisis mentality	Aftershock risk
		Duration of shaking
Frequency	Is it constant through time?	Is it constant through time? Stress-dependent?
Magnitude	Volcanic Explosivity Index	Moment Magnitude Scale
	Fluctuation with time	
Location	Lack of surface evidence? Multiple vents?	Faults not mapped?
	Areas at highest risk?	Relationship between adjacent faults?
		History of regional movement?
Area affected	e.g., Topographic controls	e.g., Geological site effects

in Caracas is very dense indeed, with poorly constructed houses almost on top of each other. Poverty, population growth, and lack of awareness have led to massively increased risk. This is a problem around the world, and one that is growing. It also occurs on multiple scales, from local to international. Tragedies such as the 2004 tsunami affect the global community, not only because citizens from around the world were impacted (for example, on the beaches in Thailand), but also because of the networks of trade agreements, NGOs, and the moral responsibility of wealthier nations to provide aid. The human geography of natural disasters is thus extremely complex. While working at the local level is of prime importance in providing resilience and education, national and international plans are required so that the broader economic and societal impacts are mitigated.

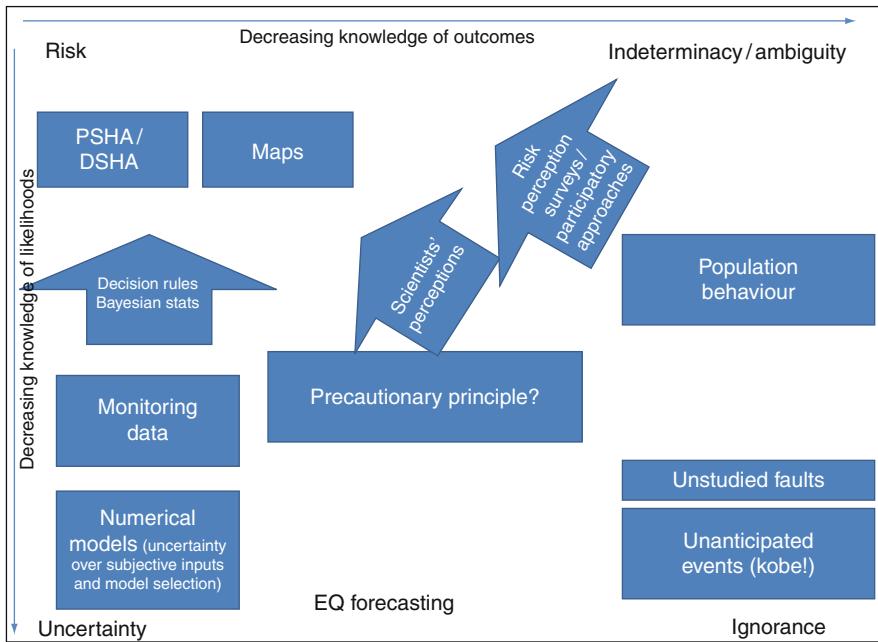
## Summary: Categorizing Uncertainty

---

This chapter deals with two rather different risk discourses: earthquakes and volcanoes. However, there are some key similarities: both suffer from the lithospheric problem (they originate in the deep earth and so are difficult to analyze from the surface); both are very difficult to predict; scientific knowledge about each is growing but relatively youthful; and both are generally high impact but low probability, particularly at higher magnitudes. These are fundamentally issues that are embedded within epistemology, philosophy, and sociology, but seldom discussed in this context. The chapter will therefore begin by discussing the history of each of these risks, how they have been dealt with by the scientific community and how publics and governments have responded to the risk. It will then discuss current themes in risk assessment and management research for each risk, and finally suggest some future directions.

Much of the information available about geophysical risks comes from the scientific community. This means that the role of scientists in crisis management can be very important, and therefore that there is a lot of pressure on people who are not trained as policy advisors. The social context of scientific data gathering, modeling and use is therefore fundamental in understanding the social role that scientists have, and the extent of their accountability – which has been questioned in recent years. This chapter will therefore seek to provide a highly interdisciplinary discussion of some of the key issues in the management of risks from geophysical hazards. It is not intended to focus on scientific methods and models: there is a huge literature on this subject. Rather, it examines the social context both of science, and of the hazards themselves, proposing a holistic and context-based approach to understanding risk and uncertainty.

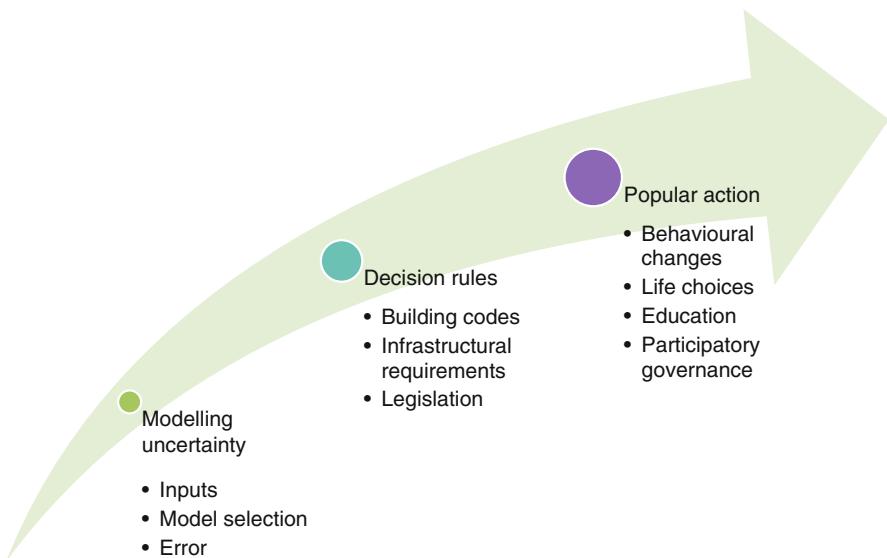
In order to provide further structure to this vast topic, Fig. 14.4 shows a potential framework for use in thinking through the transdisciplinary context of the geophysical hazards. It is tailored to earthquakes; a volcanic version is given in Donovan (2010), and the framework is based on Stirling (2007) and Wynne (1992), with some alterations for the specific case of geophysical hazards. It is a useful framework because it distinguishes between different types of uncertainty. Risk refers to uncertainty that is quantifiable and identifiable: a specific threat for which a probability can be calculated – and this definition is more focused than the traditional hazard times vulnerability formulation, though that remains conceptually helpful. It is in this part of the diagram that most risk assessments take place. However, Stirling (2008) argues that many attempts to manage uncertainty “close down” to risk, ignoring the other three aspects of



**Fig. 14.4**  
Contextualizing the issues in seismic risk research

the uncertainty, or smoothing them over. In this framework, “uncertainty” applies to situations where the outcome is known, but a probability cannot be put on it. This might be true, for example, for nonlinear numerical models, where there is uncertainty about input parameters and model selection, and about how the results of the modeling can be translated to risk assessment. “Indeterminacy” (or, for Stirling (2007), “ambiguity”) refers to situations that can be dealt with quantitatively in theory, but the outcome is unknown. This most often refers to social behaviors in the present context – understanding how populations and policymakers, and indeed scientists, will respond to events and information. Finally, “ignorance” is a very important aspect of geophysical risk assessment. It refers to the “unknown unknowns,” those things that are not anticipated. There are multiple examples of these in the recent history of the geophysical hazards, such as new phenomena observed on Montserrat, or the failure to anticipate seismic risk in particular areas.

The high level of uncertainty inherent in predicting the geophysical hazards has produced several major challenges for those seeking to manage and mitigate them. The uncertainty of the scientific methods coupled with the procedures involved in mitigating these risks and the need to involve populations in the preparation leads to a snowballing of uncertainty and indeterminacy from the scientific domain through the policy domain and into the wider public. This is illustrated in [Fig. 14.5](#). One response to this from the academic community has been to carry out risk perception and communication surveys. However, this is one area where qualitative methods are very important, and where philosophy and sociology have a useful role to play.



**Fig. 14.5**  
The snowballing-uncertainty effect

## History

The history of “geophysical risk” as a concept is relatively recent: the understanding of geological hazards and their sources has developed exponentially in the last 200 years, as their causes have been analyzed with increasingly accurate and reliable instruments. In part, this relates to technological and scientific development in other areas – the prediction and awareness of geophysical hazards is in some respects a luxury of the modern world, and is very dependent on plate tectonic theory, which was introduced by Alfred Wegener in 1912, but did not become widely accepted until the 1950s and 1960s. The definition of the theory is generally attributed to McKenzie and Parker’s 1967 *Nature* paper and Morgan’s 1968 essay. Thus, the cornerstone of modern geophysical theory is very young (Oreskes and Le Grand 2003), and scientific methods for earthquake and eruption prediction are correspondingly youthful. There are therefore several challenges in documenting the historical context of volcanic and seismic risk – volcanology and seismology as scientific disciplines have discrete histories, and it is not entirely clear at what point the idea of risk in relation to these endeavors really took hold. In general, it was reactive, and this chapter suggests that current research in both of these disciplines has been dominated in many ways by past earthquakes and eruptions, rather than pure “blue skies” research: Individual events have played key roles in the progression of understanding – both of the physical processes and of the nature and concept of geophysical risk.

Earthquakes and eruptions themselves have of course occurred throughout the earth’s history, and have been associated with mass extinctions and the collapse of civilizations (Francis and Oppenheimer 2004). The broader recognition of these possibilities is however relatively modern – those affected by the AD 79 eruption of Vesuvio, for example, were unaware of the risk prior to the eruption. Indeed, social acceptance of volcanic and seismic risk is

currently a major project in many parts of the world – one that has unfortunately postdated much infrastructural development. Both volcanic and seismic risks have been relatively slow to penetrate the social consciousness, in spite of their long histories. This is partly a function of their geographical distribution and of the long return periods of events in a specific area.

This section will outline the recent history of geophysical risk awareness, using examples to highlight some of the major issues in risk assessment and management. It is important to note at the outset, however, that risk assessment in seismology and volcanology is extremely unusual, though increasing. Much of the focus to date has been on hazard assessment, in recognition that risk assessment requires further (uncertain) input about populations, infrastructure, and mitigation measures (e.g., Wang 2009). However, risk assessments are increasingly being demanded by governments, and the scientific community is attempting to bridge the gap between what science is capable of doing, and the results that are required for policy making. Some examples of this will be given under “Current research,” below. First, however, the recent history of seismic and volcanic risk will be discussed.

## Earthquake Risk

---

It could be argued that the first “modern” earthquake in the West was the 1755 Lisbon earthquake, which was the largest documented event to affect Western Europe (Chester 2001), and which generated a tsunami in the Atlantic. This event generated a renewal of interest in seismology in the eighteenth century, and also precipitated the development and use of earthquake-resistant building techniques. Further devastating earthquakes occurred in Calabria in 1783, and then in Chile in 1835 (for a full list of large earthquakes, see [Table 14.3](#) and <http://earthquake.usgs.gov/earthquakes/world/historical.php>). Following the famous 1906 San Francisco earthquake, Henry Reid observed that earthquakes are the response of the crust to the buildup of stress on fault-lines (“elastic rebound theory”). Faults fracture as a result of stress buildup, and the fracture may propagate long distances along the fault depending on the energy of the earthquake. The 1906 earthquake has become notorious for its societal impact, too: fires burned in the city for 3 days, 3,000 people were killed and 225,000 left homeless. As a result of this event, scientific knowledge about earthquakes increased dramatically – and the San Andreas fault became one of the most feared in the world.

In 1960, a magnitude 9.5 earthquake – the largest ever recorded – occurred in Chile, sparking a Pacific-wide tsunami that killed people as far away as Hawaii, Japan and the Philippines. Four years later, in 1964, a magnitude 9.2 earthquake occurred in Alaska, again causing a tsunami that killed over a hundred people on the US West Coast. These events marked a renewed interest in the prediction of earthquakes and the assessment of seismic hazard. In the late 1960s, the Probabilistic Seismic Hazard Assessment procedure was devised in Mexico and the USA, based on modeling earthquakes as Poisson processes (McGuire 2008). It also has a strong history of Bayesian methods – used by Esteva (1969) to facilitate decision-making – thus combining frequentist and subjective probabilistic methods. The emphasis, however, is on the magnitude and frequency distributions of the earthquakes on a particular fault – a fact that assumes that this in some way represents the physical parameters of the system. Complexities may arise as a result, for example, of multiple, interlocking fault systems whose stress distribution affects one another: as knowledge about the system increases, epistemic uncertainty can also increase because scientists become aware of more unknowns in the natural system.

**Table 14.3**  
**Significant earthquakes in the historical record**

Date	Location	Magnitude	Deaths
856 AD	Damghan, Iran	Unknown	200,000
1138	Aleppo, Syria	Unknown	230,000
1556	Shensi, China	8.0	830,000
1693	Sicily, Italy	7.5	60,000
1700	Cascadia	9.0	N/A
1755	Lisbon, Portugal	8.7	70,000
1783	Calabria, Italy	Unknown	50,000
1811	New Madrid, USA	7.7	Several
1819	Gujarat, India	Unknown	2,000
1838	San Francisco	6.8	
1843	Leeward Islands	8.3	5,000
1857	Naples, Italy	6.9	11,000
1868	Big Island, Hawaii	7.9	77
1868	Arica, Peru (now Chile)	9.0	25,000
1875	Northern Colombia	7.3	16,000
1877	Offshore Tarapaca, Chile	8.3	34
1892	Imperial Valley, California	7.8	
1899	Cape Yakataga, Alaska	8.0	
1899	Menderes Valley, Turkey	6.9	1,100
1902	Guatemala	7.5	2,000
1903	Turkey	7.0	3,500
1905	Kangra, India	7.5	19,000
1906	Offshore Esmeraldas, Ecuador	8.8	1,000
1906	San Francisco, California	7.8	3,000
1907	Qaratog, Tajikistan	8.0	12,000
1908	Messina, Italy	7.2	70,000
1915	Avezzano, Italy	7.0	32,610
1920	Ningxia, China	7.8	200,000
1923	Kamchatka	8.5	
1923	Kanto, Japan	7.9	143,000
1927	Tsinghai, China	7.6	40,900
1931	Xinjiang, China	8.0	10,000
1933	Sanriku, Japan	8.4	2,990
1934	India–Nepal	8.1	10,700
1935	Quetta, Pakistan	7.5	30,000
1939	Chillan, Chile	7.8	28,000

**Table 14.3 (Continued)**

Date	Location	Magnitude	Deaths
1939	Erzincan, Turkey	7.8	32,700
1948	Ashgabat, Turkmenistan	7.3	110,000
1949	Khait, Tajikistan	7.5	12,000
1952	Kamchatka	9.0	
1957	Andreanof Islands, Alaska	8.6	
1960	Agadir, Morocco	5.7	10,000
1960	Chile	9.5	1,655
1962	Qazvin, Iran	7.1	12,225
1964	Prince William Sound, Alaska	9.2	128
1968	Dasht-e Bayaz, Iran	7.3	12,000
1970	Yunnan Province, China	7.5	10,000
1970	Chimbote, Peru	7.9	66,000
1974	China	6.8	20,000
1976	Guatemala	7.5	23,000
1976	Tangshan, China	7.5	255,000
1978	Iran	7.8	15,000
1985	Michoacan, Mexico	8.0	9,500
1988	Spitak, Armenia	6.8	25,000
1989	Loma Prieta, California	6.9	63
1990	Western Iran	7.4	50,000
1994	Northridge, California	6.7	60
1995	Kobe, Japan	6.9	5,502
1999	Izmit, Turkey	7.6	17,118
2001	Gujarat, India	7.6	20,023
2003	Bam, Iran	6.6	31,000
2004	Sumatra-Andaman, Indian Ocean	9.1	227,898
2005	Northern Sumatra, Indonesia	8.6	1,313
2005	Pakistan	7.6	86,000
2007	Southern Sumatra, Indonesia	8.5	25
2008	Eastern Sichuan, China	7.9	87,587
2009	L'Aquila, Central Italy	6.3	295
2010	Haiti	7.0	222,570
2010	Offshore Bio-Bio, Chile	8.8	577
2011	Near east coast of Honshu, Japan	9.0	13,538+ (14,000 missing)

The selection here is based on magnitude, fatalities, and significance in the scientific debates. This list is only a small fraction of the available data. As the historical record improves exponentially in the last few decades, the distribution here cannot be taken as representative. Data is taken from the USGS. Figures for the Japan 2011 earthquake are correct at time of writing (April 2011)

The recent history of earthquake activity demonstrates the challenges of deciding whether or not to attempt to predict events. In 1983, the USGS predicted that there would be an earthquake in the Parkfield area within 5 years of 1988. It actually occurred in 2004 – after the window had expired. In 1989, the Loma Prieta earthquake occurred (e.g., Reasenberg and Simpson 1992). Loma Prieta had been associated with a seismic gap (an area of fault that had not moved for some time), but a prediction was not made. Both of these events led to public questions about the scientists' credibility and responsibility. Then, in 1995, the Kobe earthquake occurred in Japan. Kobe had not been recognized as an area of high seismic risk, and there was a level of complacency in Japan about its ability to cope with earthquakes – perhaps because there had not been a major earthquake for many years (Wisner et al. 2004). The devastating impact of the Kobe earthquake, killing over 5,000 people and destroying 136,000 buildings, came as a huge shock to Japanese infrastructure. It resulted in the shutdown and replacement of Japan's "earthquake prediction program," and an extensive national disaster preparedness campaign. In recent years, disaster managers from Kobe have been involved in assisting other countries as they recover from earthquakes, such as the Iranian response to the 2003 Bam earthquake, demonstrating the usefulness of global knowledge economies in managing geophysical hazards.

The last decade (2000–2010) has witnessed several major earthquakes. Two of these were particularly devastating in terms of casualties: the Sumatra-Andaman earthquake in 2004, and the Haiti earthquake in 2010 (► *Table 14.3*). The 2004 earthquake was very large – 9.2 on the moment magnitude scale – and caused considerable seafloor deformation, which generated a tsunami. This was responsible for most of the casualties, and the death toll could have been significantly reduced by provision of early warning (Sieh 2006; Kanamori 2006). The absence of a tsunami warning system in the Indian Ocean – in spite of there being one in the Pacific – was partly the result of social and economic inequalities, and partly a lack of recent memory of tsunami in this region (though in 1883 the eruption of Krakatau caused a tsunami; see below). The global distribution of wealth can have a very large impact on the ability to respond to geophysical risk – and many of the risks are higher in the developing world (Wisner 2004).

The 2010 Haiti earthquake, by contrast, was a moderate-sized event – 7.0 on the moment magnitude scale. Its destructive power was related to its location close to Port au Prince, its relatively shallow depth (8.1 km), and the inadequacy of local infrastructure to withstand the shaking. Buildings in Haiti had not been built in accordance with earthquake building codes, and thus collapsed due to the effects of ground shaking. This was exacerbated by several strong aftershocks, and the soft geology of the region (which amplified the shaking). Thus, a lack of preparedness and awareness of the risk were major factors in the high death toll (222,570). This demonstrates the importance of an integrated approach to disaster risk reduction and management.

On March 11, 2011, a magnitude-9 earthquake struck Honshu, Japan, and caused a Pacific-wide tsunami, killing people around the Pacific Rim, and devastating large areas of the Japanese coast. Disruption to the power supply of a nuclear plant in Fukushima caused a Level-7 nuclear crisis – the highest level on the international scale. In part, this was due to the age of the plant and the underestimation of likely tsunami height when defenses were constructed. This incident demonstrated the vulnerability of the most developed nations to natural disasters: In terms of earthquake-preparedness and engineering, Japan is arguably the most advanced nation in the world. In this case, technological advancement both decreased vulnerability and

increased it – while buildings in Tokyo were able to withstand substantial shaking, the presence of nuclear facilities and a heavy dependence on power supplies significantly increased the risk. Additionally, the hazard itself was of a higher magnitude than defenses on the coast had been designed to withstand. The combination of earthquake and tsunami in Japan did not cause the same loss of life as the similar event in Indonesia in 2004, perhaps because of engineering solutions, but it showed that there is some way to go in multihazard designs for key facilities, particularly in a technologically advanced society.

## Volcanic Risk

---

Recent archaeology has revealed the dramatic impact of volcanic eruptions on civilizations throughout history – most notably perhaps in the destruction caused by the eruption of Thera on Santorini around 1650 BC, associated with the destruction and burial of Akrotiri (Francis and Oppenheimer 2004), and also with several Mediterranean legends. The notorious eruption of Vesuvius in AD79 could be taken as the first documented volcanic eruption, described by Pliny the Younger in some detail. The extensive looting of Pompeii and Ercolano in the seventeenth and eighteenth centuries demonstrates the interest that was accorded to them, and the travels of early scientists such as William Hamilton are indicative of the growth of curiosity about the workings of the Italian volcanoes in particular. Hamilton documented the lava flows on Vesuvius, and identified the older Mount Somma as the source of the AD 79 eruption (Hamilton 1772). The Enlightenment period was crucial in the development of the geological sciences in general, and of interest in volcanology in particular – and this continued throughout the nineteenth century (for excellent accounts of the history of ideas in volcanology, see Sigurdsson 1999 and Young 2003). This is reflected too in artistic endeavors, such as the Romantic movement, which were inspired by the “sublime and terrible wonders of nature”.

The eighteenth and nineteenth centuries also have a rich history of volcanic eruptions, particularly in terms of climactic impacts – the injection of sulfate aerosols into the stratosphere can cause climate forcing, resulting in cooling and heating in different areas depending on the location of the plume. In 1783, the Lakagigar eruption in Iceland caused failed harvests across northern Europe (Witham and Oppenheimer 2004). Then, in 1815, Tambora volcano in Indonesia erupted. This was the largest eruption in historical time, and caused spectacular sunsets around the world, as well as altering the climate (Zerefos et al. 2007). The early nineteenth-century eruptions of Vesuvius prompted the founding of the Osservatorio Vesuviano in 1841 to monitor the volcano, as growing numbers of tourists were visiting it.

Several major eruptions have had significant impacts on social and scientific awareness of volcanic risk. An important example was the 1883 eruption of Krakatau in Indonesia, which caused a tsunami in the Indian Ocean that killed 36,000 people. It has been argued that this was the first natural disaster to be communicated globally at speed (Winchester 2003): It appeared in the London newspapers shortly after it occurred. Nineteen years later, Mount Pelée on Martinique in the French West Indies erupted, killing all but two of the population of St Pierre. This event is often taken as the origin of modern volcanology. It was described by Alfred Lacroix in 1904 (Lacroix 1904). In 1924, the Bulletin of Volcanology was founded, to promote academic research in the subject. Researchers observed and began to analyze precursory signs, such as earthquakes prior to eruptions (e.g., Finch 1943; MacGregor 1949) and changes in ground tilt (e.g., Waesche 1942). As far back as 1937, R.L. Ives noted that “our modern civilisation is more easily damaged

than ever before in history” – documenting an awareness of growing societal vulnerability to both earthquakes and volcanoes (Ives 1937). Again, persistently active volcanoes were the source of a great deal of data – most notably the Italian and Hawaiian volcanoes.

A major turning point in societal awareness and scientific understanding of volcanic risk was the 1980 eruption of Mount St Helens in Washington, USA, which killed 57 people and produced a huge ash cloud that dispersed over large parts of the USA (e.g., Moore and Rice 1984). It was caught on film, and the pictures have become classics in the volcanological literature. While the eruption was in a relatively remote place, 57 deaths occurred as a result of it. Those who died had either refused to move, or had been working for the timber industry, which had put pressure on the Governor of Washington to keep access open to its workers. Prior to the eruption, there had been long-term warnings that an event was likely given the history of the volcano, and some hazard perception studies had been carried out. During the eruption, further studies of hazard perception were carried out among local residents (Perry and Greene 1983; Lindell and Perry 1993). These studies suggested that, unsurprisingly, popular perception of the risk from the volcano increased dramatically during the first few months of the eruption, and followed patterns described in other disciplines – with people being more open to information from trusted sources, for example. Lindell and Perry (1993) also discuss the transition in risk perception from the acute threat to a chronic threat (Mount St Helens continued erupting throughout the early 1980s): This is an important aspect of risk perception during eruptions, and has also been witnessed on Montserrat, where the eruption has been ongoing, episodically, for 15 years. The potential for “chronic crisis mentality” in the governance of long volcanic eruptions has been recognized by scientists working on Montserrat, and increasingly by local officials. There is thus a long-term mental shift that has to take place as procedures for risk assessment and management are established and refined. Examples of this include the management of long-term eruptions on Etna and Hawaii. Adjustments have to be made on an individual level, and on a societal level. Ongoing volcanic eruptions can have significant impacts on national identity and human geography.

The largest eruption of the twentieth century occurred in 1991 at Mount Pinatubo in the Philippines (Newhall and Punongbayan 1996). This eruption was regarded in many ways as a successful management of volcanic risk, because thousands of people were evacuated and many lives were saved. Scientists in this case were able to gather sufficient data to be convinced that an eruption was very likely, and eventually managed to convince local officials – and, via lobbying the Pentagon, the nearby US air base – that an evacuation was necessary. However, volcanic mudflows continue to affect the area many years later, even after the end of the eruption.

Other important events in the recent history of volcanology include the 1993 eruption of Galeras volcano in Colombia and the eruption of Mount Unzen in Japan – both of these involved the deaths of volcanologists, demonstrating the uncertainties not only about eruption prediction, but also about the extent and capabilities of volcanic hazards such as pyroclastic flows. This was highlighted on Montserrat in 1997, when 19 people were killed (in the exclusion zone) by highly mobile pyroclastic flows. In this case, scientists were almost killed, and the microzonation of the island was regarded with hindsight as extremely dangerous (Aspinall et al. 2002): geographical definition of the hazardous areas was subject to high errors, partly because the mobility of pyroclastic flows was poorly understood. The eruption allowed pyroclastic flows to be studied and modeled in detail.

It would be remiss to conclude this section without a brief discussion of the 2010 eruption of Eyjafjallajökull in South Iceland – a small eruption that brought aviation in Northern

Europe to a standstill because of the risk to aircraft from the volcanic plume. It is a fascinating example of the problems of geophysical governance. As was reported ad nauseam during the “ash crisis,” the threat to aircraft from volcanic eruptions has been known for decades. In February 2010, for example, an explosion at the Soufriere Hills Volcano on Montserrat suspended air traffic in the Eastern Caribbean and closed the international airport in Antigua. The lack of preparedness in Northern Europe was largely the result of industrial and political apathy: The event was not a surprise to the scientific community. This emphasizes a further point about the nature of geophysical risks: The scientific community and the lay community may differ greatly in their view about the importance of a particular threat, and this renders any definition of risk thresholds problematic.

Thus, in spite of the long history of volcanic crises, in 2009 Louisiana Governor Bobby Jindal criticized Obama’s government for funding “something called volcano monitoring”: The recognition that volcanoes pose a serious threat to human welfare is still not widely accepted even in the most developed parts of the world. When the volcano on Montserrat began erupting in 1995, for example, very few of the islanders even knew that it was a volcano. New volcanic phenomena are still being observed and documented, and there are types of volcanic eruption that have not been observed in recorded history – such as so-called supervolcanic eruptions. There is a frequency–magnitude relationship in the geological record: Larger eruptions occur more rarely than smaller ones (Mason et al. 2004). However, the nature of that relationship, its relationship to physical processes, and thus its reliability for predictive purposes is contested. For example, Katla volcano in South Iceland erupted 20 times in a 1,000-year period. It therefore erupts every 50 years. However, the last eruption was in 1918; is it therefore meaningful for scientists to argue that it is “overdue”? Or has the magmatic system shifted in some way? This tension between statistical and physical models will be discussed below.

## Summary

---

This section has discussed some of the defining moments in the history of volcanic and seismic risk. It has shown some of the key ideas that have emerged in the last century, as well as the challenges. In particular, the importance of specific events has been emphasized, both for producing data and scientific advancement, and for raising social awareness of risk. In earthquake hazard, the scientific challenges that have arisen from recent events concern the nature and appropriateness of earthquake prediction, which are debatable. The Kobe earthquake and the Bam earthquake both demonstrated unpreparedness in the absence of physical evidence, as well as faulty assumptions in earthquake prediction methods, for example. Societal challenges related to this include the need for public education, enforcement of building codes, and a precautionary approach. In volcanology, scientific challenges relate to multidisciplinary consensus-building, managing uncertainty in models, and preventing overconfidence in the face of the “unknown unknowns”. Socially, education remains very important, as do emergency planning and infrastructural preservation.

## Current Research

---

This section is structured around two aspects of risk research in the geophysical hazards – the technical, scientific management of uncertainty, and the social aspects of risk. This is not

intended to be dichotomous, but to provide a framework for examining the different discourses involved. While scientific research in volcanic and seismic risk does tend to “close down” toward risk (Stirling 2008) – quantification of probabilities being the goal – combination of social scientific and scientific methods via reflexive awareness and narrative management may be a powerful way forward for the physical and social sciences to work together. This will be the argument of the final section of the chapter.

## Scientific Uncertainty

---

The challenge of handling uncertainty in science has been increasingly recognized (e.g., Taleb 2007) – not least through the climate change discourse and the controversy surrounding the limitations of scientific knowledge and practices. In seismology and volcanology, very high levels of uncertainty may be associated with predictions. This uncertainty comes from a number of sources – such as model selection and inputs (e.g., Oreskes et al. 1994), historical/geological data and its gaps, expert judgment and interpretation, observational error, computational error, and instrumental error (see also Shackley and Wynne 1996; Shackley et al. 1998). A distinction that is commonly in use separates epistemic uncertainty (lack of knowledge) and aleatory uncertainty (randomness of nature; e.g., Woo 1999). The problem with this distinction is that it oversimplifies the problem somewhat – both are generally combined in the occurrence of particular events, and the distinction between them is not always clear. In addition, even models that attempt to calculate both uncertainties are not foolproof – the “unknown unknown” element is increasingly recognized (Spiegelhalter and Riesch, *in review*) as problematic, and this has been demonstrated in the recent history of volcanic and seismic crises, where previously unobserved phenomena have been documented. The development of scientific models for these phenomena is encouraging, but models are limited by their simplification of natural processes – and may be misinterpreted by the public, particularly where they look impressive (Hulme and Mahony 2010). The authority of scientific knowledge is contingent on a variety of scientific and social considerations (Fischer 2004; Hulme 2009). This section will discuss the current state of research in seismic and volcanic risk and uncertainty in turn.

## Earthquakes

Seismologists have been carrying out probabilistic seismic hazard assessment (PSHA) for over 40 years. It is based on historical seismicity in particular areas, and has been for some time the dominant method for calculating seismic hazard (e.g., by the US Senior Seismic Hazard Analysis Committee). Bayesian inference has also been applied (e.g., Faenza et al. 2010). However, in recent years, there has been a growing dissatisfaction with probabilistic methods, and new techniques – “deterministic seismic hazard assessment” (DSHA) or “neo-DSHA” – have been developed. This has led to a heated debate about the nature of earthquake prediction (e.g., Krinitzsky 1995; Bommer and Abrahamson 2006) and the most appropriate methods for representing earthquake hazard for policymakers. In part the problem is one of managing aleatory uncertainty – not included in the original formulations of PSHA in the 1960s (McGuire 2008; Bommer and Abrahamson 2006), but later added to the procedures.

However, as McGuire (2001) argues, probabilistic and deterministic approaches can be complementary and for decision-making purposes, probabilistic assessments are necessary, even though they can include deterministic methods (Bommer and Abrahamson 2006).

The nature of uncertainty in seismic hazard assessment is particularly disputed, with some authors claiming that DSHA does not take all the uncertainty into account, and others that it does (Wang 2010). A problem that is well recognized in DSHA is that it does not take return period into account – a problem that is significantly complicated by evidence that, for example, earthquakes may occur in clusters, and the measurement of return period in PSHA (usually by representing it as a Poisson process) may thus not be adequate. It is clear from this brief review that seismic hazard assessment remains a highly contested area of research. Terminology is a big problem: “Deterministic” is not always used in the sense of classical determinism, and “probabilistic” assessments may be frequentist or subjectivist. This can cause significant confusion. However, there are also authors who argue against the use of probabilities in general (e.g., Castanos and Lomnitz 2002), on the ground that they cannot be reproduced by experiment. This is a particularly interesting aspect of scientific understandings of uncertainty: the emphasis for some scientists has to be on what is verifiable and physically observable. In DSHA, this refers to all the information that it is possible to obtain for a particular faultline – “hard facts” rather than statistics (Castanos and Lomnitz 2002). These authors would dispute the use of statistical methods at all as a means of quantifying uncertainty, arguing that statistics “works best with plenty of data”.

The selection of probability distributions to represent the process of earthquake recurrence (particularly Poisson distributions) in PSHA is inevitably subjective itself: All model selection is subjective, as is the selection of data and inputs used in the assessment. There is no way of knowing whether or not all the information has been taken into account – as the cluster problem shows. Even in deterministic methods, there are possibilities that geological information has been misinterpreted, that instruments have been inaccurate, and so on. It is therefore difficult to argue that any earthquake hazard assessment can be truly objective – as pointed out by several authors. However, those that make extensive use of expert judgment have been particularly vulnerable to attack. Expert elicitation, for example, involves quantification of data using experts as processors, and is therefore highly subjective, rendering it open to accusations of being unscientific and inappropriate. At the same time, it allows the translation of scientific data into risk assessments, bridging the gap between science and the need to advise policymakers.

Different theories about the recurrence rates of earthquakes have been put forward, and some faults appear to corroborate particular models. However, there is no foolproof explanation for the behavior of faults in general. Thus, some faults seem to have “characteristic earthquakes” of a particular magnitude that recur at relatively predictable intervals, but others do not. Other faults exhibit clustering of earthquakes in time. The nonlinear behavior of faults requires detailed mapping studies to be carried out. However, geological mapping can be a matter of interpretation, and this requires particular care in managing uncertainty: For example, distinguishing a large earthquake from a series of small earthquakes that occurred in close temporal proximity can be challenging in the field. Thus, there are a variety of conceptual and quantitative challenges in understanding the physics of earthquakes. Nevertheless, the existence of relatively robust probabilistic maps is a necessity for planning, and has been applied in many key areas (e.g., Stein et al. 2006).

Uncertainties also arise in the expression of the outcomes from seismic hazard assessment: There are several possible variables that can be estimated and different ways of expressing the same variable (Wang 2010). Thus, for example, the magnitude of a particular earthquake could be stated in terms of its moment magnitude, or its likely peak ground acceleration. Similarly, the temporal spread of earthquakes may be expressed in terms of return period (how often, on average, an earthquake of that magnitude occurs in a particular place), or as the “probability of exceedance” for particular degrees of ground shaking in a particular time period. As well as expressing magnitude and frequency, the likely spatial distribution of shaking has to be expressed, and this is commonly done using maps. Unfortunately, the ability of members of the public – and policymakers – to understand hazard and risk maps is not clear (Newhall et al. 1999; Haynes et al. 2007). This is a further area for social and physical scientists to work on together.

Seismic risk, to take the argument a step further, involves a much greater range of variables, including potential economic losses and fatalities. This is complicated by the involvement of multiple parties in assessing and mitigating risk, not least because the language used by earthquake engineers often differs from that used by seismologists – and that used by disaster managers may well be different again! Spence (2009) details some of the challenges currently facing the seismological and earthquake engineering communities, particularly with respect to building code implementation around the world. He notes the importance of international collaborative ventures in meeting some of these challenges. In particular, the need to estimate losses can be problematic because obtaining data about buildings and infrastructure is not straightforward and may be politically sensitive. Efforts are underway, however, to facilitate loss modeling on a global scale (Spence 2009).

## Volcanoes

Volcanology is a smaller discipline than seismology, and while tremendous advances are now being made, they have perhaps been slower to occur – perhaps because of the relative paucity of volcanic eruptions and hence observational data. In addition, there are complexities induced by the nature of volcanism – the need to understand seismicity, ground deformation, gas emissions, gravitational variation, electromagnetic signals, hydrochemistry, and other types of signals, for example, some of which might appear to give contradictory indications of whether or not a particular volcano will erupt. Most volcanoes, however, are not adequately monitored (Ewert and Newhall 2004), or not monitored at all: This may render any kind of prediction for a particular location very challenging because there is no baseline data and the first discernable activity may be explosions from the crater (Tilling 2008; Lowenstern et al. 2006). Monitoring networks allow the detection of migrating magma in the volcanic system; this data has to be interpreted in order to ascertain how likely it is that the magma will reach the surface. This has been increasingly attempted using subjective probabilistic methods, particularly as the presence of aleatory uncertainty in the natural system is recognized by scientists. The main methods currently in use are Bayesian Event Trees (e.g., Newhall and Hoblitt 2002; Marzocchi et al. 2006, 2007) and Belief Networks (Aspinall et al. 2003; Hincks 2005), and structured expert elicitation (Cooke 1991; Aspinall 2006, 2010; see also O’Hagan et al. 2006). However, these methods are by no means universally accepted in the

scientific community (Donovan 2010). Many other methods for loss modeling, for example, can be controversial because of the need to value human lives (e.g., Marzocchi and Woo 2009).

Bayesian methods are a challenge to determinists because of the selection of the prior, which is subjective (Gelman et al. 2004; Gelman 2008) and a matter of interpretation. However, many scientists support the use of these methods in recognition of the need to display the available information logically, and to produce probabilistic risk assessments. Probabilities to populate event trees may be calculated using frequentist priors – based on the historical record of volcanism in a particular place – or using subjective methods such as expert elicitation. Elicitation is gaining popularity as an effective way to quantify risk assessments using all the available data, but again faces criticism from scientists who are concerned about subjective methods: “Evidence-based volcanology” (Aspinall et al. 2003) faces many of the criticisms faced by its predecessor, “Evidence-based medicine” (e.g., Sackett and Rosenberg 1995; Sackett et al. 1996). Deterministic models have also been produced to forecast volcanic eruptions (e.g., Voight 1988; Kilburn 2003; Kilburn and Sammonds 2005; Melnik and Sparks 1999). As scientific knowledge about volcanoes increases, and further data is gathered from volcanic systems around the world, possibilities arise for more sophisticated modeling of global volcanism – perhaps using ensemble models. This is the aim of several large projects that are currently gathering data of different kinds and developing cybertools for its analysis. Some multidisciplinary projects in recent years have generated some very practical and important resources for emergency planning. One example was the Exploris project (Baxter et al. 2008), which examined four European volcanoes and generated probabilistic models for eruption scenarios (e.g., at Vesuvius; Neri et al. 2008; Spence et al. 2008; Zuccaro et al. 2008).

The importance of the social context of scientific activities in volcanology has been commented upon in the literature since the failed eruption of La Soufrière de Guadeloupe in 1976, where a dispute between leading volcanologists became public and undermined the scientific effort to monitor the eruption (Tazieff 1977; Fiske 1984). It was partially in response to the problem of achieving consensus that the expert elicitation procedure now used regularly onMontserrat was introduced in 1995, and these issues also led to the production in 1999 of a set of guidelines for the conduct of scientists during volcanic crises (Newhall 1999). This was a significant development because the guidelines highlighted the impact of the social context of science on scientists’ ability to provide advice. They also emphasized the potential for tension between those with social responsibilities for monitoring volcanoes and those who wish to carry out research (Donovan 2010).

## Social Aspects of Risk and Uncertainty

### Risk Perception and Communication

As discussed in the previous section, social uncertainty may be a major factor in scientific uncertainty, in that different scientists have different ideas about what uncertainty itself is, what the role of models is, and how the natural world works. However, the management of uncertainty in the broader population – and indeed the communication of risk to the public and policymakers (e.g., Solana et al. 2008; Barclay et al. 2008) – provides a wealth of further challenges. In the geophysical hazards, this may be compounded by a lack of interest in the

events because of the challenges of daily living (Gaillard 2008). It is also complicated significantly by cultural variation in the perception of risk, which can be very strongly influenced by worldview and cognition. The psychology of risk perception has discussed a variety of heuristics that are applied by people in the presence of uncertainty (e.g., Slovic 2000; Slovic et al. 2004; Sjoberg 2000, 2008; Pidgeon et al. 2003).

On many volcanoes – including Mount St Helens, Pinatubo, and the Soufriere Hills – risk perception studies have shown that the public have a good understanding of the risk from the volcanoes (Gaillard 2008; Haynes et al. 2007, 2008; Bird et al. 2009; Barberi et al. 2008; Perry and Lindell 2008). They may even have a good understanding of some of the science, and, as mentioned previously, its limitations. However, this does not necessarily lead residents to make the decisions that researchers (or indeed governments) feel that they should make. Individual belief systems play a very important role in controlling behavior, as do relationships, trust, and potential gains. Handling low-probability, high-impact risks raises a variety of philosophical questions: Some scientists, for example, feel that the catastrophic risk potential from volcanoes means that even low risks should be taken more seriously, whereas the public may be more willing to take a risk, particularly if it is convenient or likely to result in gain. This question of “acceptable risk” – if such a thing exists – becomes extremely important and yet intangible.

It is in this context that work by Roeser (2006, 2007, 2009, 2010) is particularly helpful. Moral emotional judgments play a very important role in managing volcanic and seismic crises, because of the high stakes involved. Roeser (2009) argues that the apparent dichotomy between rational and emotional judgments about risk (“dual process theory”) is a false distinction: It is also a distinction that is at the heart of ideas about “acceptable risk”. There is an implied moral element here, one that has been witnessed in debates on Montserrat about the legality and morality of mandatory evacuations. In this particular case, there are two competing moral judgments. The first states that it is wrong to force people to leave their property (and in many situations this includes their livelihoods, particularly in farming communities). This judgment also appeals to ideas of the right to ownership and to take personal responsibility for one’s safety. The second judgment states that if the risk of a catastrophic event is considered to be high, then people should be moved even if this is against their will. The reasoning behind this may vary – it may be based on high levels of uncertainty in the risk assessment, felt or legal accountability of local authorities (as is the case on Montserrat), or the recognition that if a hazardous event occurs it will be too dangerous to send in rescuers – a further moral judgment. In several recent natural disasters, including Hurricane Katrina, mandatory evacuation orders were challenged by residents who wanted to take responsibility for themselves regardless of the risk of death. On Montserrat, the Governor was sued by local residents for forcing them out of their homes for several months during a period of high volcanic activity (Donovan 2010). This case also involved testimony from scientists involved – and several suffered socially as a result of the tensions on the island, because the political decisions were made based on scientific advice. The fact at the time was that the uncertainty was very high and the Governor of the island is personally responsible, under UK law, for the safety of the population.

There are cultural differences in the way that people view the morality of risk judgments. Many of those who objected to the evacuations on Montserrat were from the USA, where there is a more prevalent culture of litigation and of rights to property. There was also a colonial element – a sensitivity to being told what to do by the UK. However, in other countries, objections may vary. For example, the recent eruption of Mount Merapi in Indonesia involved

the evacuation of hundreds of thousands of people. In this case, the refusal to move was based on farming livelihoods and religious beliefs (see also Wilson et al. 2009). However, the vast majority of people evacuated because they could see that the volcano posed a significant threat. The problem on Montserrat – and in other evacuations – was that the threat was less obvious and not realized to the full extent possible. Communicating the fact that in conditions of high uncertainty, an evacuation may be justified even if the feared event does not occur is a major challenge in the management of geophysical hazards – particularly volcanoes, where such evacuations can last for months. Judgments about these issues are highly moralized for many stakeholders – and this is not a purely rational morality. It can be influenced by affect and cognitive biases, and this is at the heart of the fabled notion of “acceptable risk.” It is also the reason that this idea is so disputed, particularly among scientists.

Recent work by Dan Kahan (Kahan et al. 2009; Kahan 2010) has highlighted the role of cultural cognition in risk communication. This broadly reflects studies of the cultural perception of risk, such as those on Mount Merapi (Dove 2008), Montserrat (Haynes et al. 2008; Donovan 2010), and Etna (Chester et al. 2008). In these cases, religion played a very important role in the interpretation of the volcanic events – perhaps suggesting that a “risk society” mentality is in different stages of development in these areas, even as the global risk society looks in. Closely related is the cultural perception of uncertainty. This differs from the perception of risk in that uncertainty is often much less well defined. There can be a significant risk of a volcanic eruption, but considerable uncertainty as to its nature, its magnitude, its acceptability, the responsibility of local officials, population distribution, and so on. Some aspects of this may be quantifiable, others not. Cultural cognition suggests that the ways in which people handle risk and uncertainty are likely to be affected by their personal beliefs and value systems. The formation of groups and networks with similar values is a symptom of this (Beck 1992; Kahan 2010); a key example would be the climate skeptics, who reject the science behind climate change. Cultural cognition tends to relate to authority, equality, community, and individualism (Kahan 2010); these values have important implications for the management of volcanic and seismic crises, and their influence is traceable in the legality and morality discourses noted above. In volcanic crises in particular, the question of authority is extremely important, as is that of national identity (Donovan et al. 2011). The respect not only for local officials involved in decision making, but also for scientists providing advice, is extremely influential in whether or not people will obey an evacuation order. There is a tension between educating a population that there is a high level of uncertainty, and undermining their trust in science – unless uncertainty is adequately explained and represented in the communication of scientific advice (e.g., Hillerbrand and Ghil 2008).

In the case of volcanoes in particular, tourism is an increasing factor in increasing vulnerability (e.g., Bird et al. 2010; Leonard et al. 2008; Erfurt-Cooper and Cooper 2010). On Mount Etna, for example, thousands of tourists ascend the volcano every day in the summer. Educating tourists is rather different from educating resident populations: There is a transience about being a tourist, and the risk is clearly much lower if relatively little time is spent on the volcano. Nevertheless, this can be a major challenge in some areas. On Hekla volcano in Iceland, for example, scientists are relatively confident that they will have a warning around 20–30 min before the volcano erupts. However, Hekla is a popular hiking route in summer and takes around 4 h to climb. Therefore, a logistical problem arises: How to communicate a warning given that people may not have time to escape in any case. This itself may have cultural implications, too: Some nations might be more risk averse and keep people

away at all times, while others would prefer to take the risk in order to maximize the appreciation of natural beauty – and, perhaps most notably on Etna, the economic value of tourism.

## Risk Governance and Policy Making

Several studies of public participation have been carried out on active volcanoes (e.g., Mitchell 2006; Cronin et al. 2004). Donovan (2010) argued that it can be viewed as a form of political empathy: the consultation of local populations in the drawing up of evacuation plans is an opportunity to ensure that they feel listened to, and also to increase their knowledge and perception of the risk. However, the fact that much such consultation has in fact been carried out by researchers and not by governments highlights again the problem of governing low-probability, high-impact risks. The selection of field sites by researchers may not be representative of global risk or societal need. In general, participatory exercises have been well received by local communities. In crisis situations, though, participation cannot realistically be organized. Thus, a prerequisite of participatory governance for natural hazards is that it is preparatory not responsive. The exception would be the management of ongoing volcanic eruptions where particular issues can be dealt with at public meetings.

Recent discussions in scientific governance have concerned the use of the precautionary principle as a key aspect of governing under uncertainty. This adds another spin to the discussion about “acceptable risk”: If a risk is very low, is it reasonable to make decisions based on precaution if the decision will involve major ramifications for stakeholders? This is further complicated in the case of catastrophe risk, where the moral imperative would seem to be on preventing disaster. The problem is that the economic or societal impact of an evacuation may also be disastrous – and the resulting decisions therefore have to weigh lives against livelihoods. The nature of geophysical hazards can be very challenging for policymakers, particularly given the low probability of their occurrence and the paucity of votes involved in preparing for events that have not happened in living memory.

## Summary

This section has attempted to document a very wide range of disciplines in relation to risk research in the geophysical hazards, spanning both social and physical sciences. Risk unites disciplines, but also draws attention to the methodological and philosophical differences between them. In particular, the differences between the epistemology of tradition research (sometimes referred to as “mode 1” science) and research that is driven by societal need (“mode 2”) may be significant, but also blurred: Individual scientists are motivated by different factors, and have differing beliefs about the nature of science and its role in political decision-making. This is also true of social scientists. Coordination of multidisciplinary projects is therefore an important challenge in the current research climate.

Recent history suggests that the number of people adversely affected by volcanic eruptions and earthquakes will continue to increase, and that the impacts will be severe (Huppert and Sparks 2006; Jackson 2006). New research into both scientific and societal methods for prediction and management of geophysical hazards is crucial for protecting populations.

This includes hazard modeling and forecasting, education of policymakers and publics, engineering solutions, and social and political action.

## Further Research

---

This section follows from the previous two sections, and summarizes potential areas for expansion in research in geophysical risks. It deals both with scientific challenges and social scientific challenges, framing them in a philosophical and epistemological context.

### Uncertain Science

---

Both volcanological and seismological communities are currently engaged in a philosophical and epistemological struggle to manage uncertainty in the assessment of risk. An important first step in this process is to define the terms, which have become very confused in both fields. Secondly, it would be helpful to examine the social aspects of the scientific assessments, and particularly the role of expert judgment in providing scientific advice under high levels of uncertainty. This is in part a philosophical and epistemological problem that requires the exploration of new ways of doing science in the social context. Sociologists of Scientific Knowledge have delineated “mode 1” and “mode 2” science as traditional and applied science, respectively (Gibbons et al. 1994). However, this distinction breaks down where science enters the policy sphere in the geophysical hazards, as scientists attempt to retain their traditional methods in order to answer questions that science itself cannot yet answer. Risk assessments routinely challenge the nature of science and ask it to extend its epistemological basis. This has caused some confusion, and a lot of debate, among scientists, as is discussed above.

The management of scientific uncertainty in modeling has been under scrutiny in the climate change debates. This is an extremely important area of research for assessing geophysical hazards. Several types of model are relevant – global models, local models, hazard-specific models, multidisciplinary models – all of which require development. The management and understanding of the different types of uncertainty involved in modeling is also in need of further discussion. Similarly, the use of different kinds of monitoring data in models of particular volcanic systems is worthy of further research, especially in relation to the management of error and uncertainty, which varies extensively between different types of monitoring. Finally, the ongoing acquisition of new data about faultlines and volcanoes is crucial in mapping risks, and in informing models and statistical methods.

The use of subjective probabilistic methods for both hazard and risk assessment is also in need of further review and discussion. This is an important area of interdisciplinarity between social and physical scientists, and the former could take a more proactive role in understanding the social and philosophical implications of particular methods (e.g., De Finetti 1974; Corvalyn 2008), as well as in identifying social factors that could be incorporated into mitigation models. While quantification is necessary for risk management, qualitative methods from the social sciences are important in framing the use of quantitative data and its presentation. Philosophical ideas are also important in understanding the ways in which science is used, progresses, and interacts with the social and political domains. The question of how models, interpretations, and reports handle scientific and societal uncertainty is also a key epistemological challenge for

both physical and social scientists – it relates to the nature of science and its social role. This is a growing area of research in the geophysical hazards, and one with a lot of scope for further development.

## Uncertain Societies

---

Social understanding of risk from geophysical hazards has been shown to be restricted by the low probability of these events, and the consequent focus on everyday issues. Recent research on moral emotions and cultural cognition suggests that new pathways have to be found to represent risk in socially and culturally appropriate ways (Kahan et al. 2009). This requires cultural studies using qualitative methods alongside studies in expression and communication. In particular, the potential for catastrophic impacts as a result of eruptions or earthquakes requires a sensitive approach combining legal, political, cultural, and scientific perspectives to draw up plans that include moral and emotional awareness – for example, concerning attachment to property and livelihoods. In this respect, there are many studies that can be used to illustrate the importance of these aspects in the past, and the power of anecdote in this context should not be underestimated. Personal and societal narratives are a rich source of information about how particular societies handle uncertainty. Chester (2005) demonstrated the importance of religion in understanding the impact of disasters, and in planning for particular events: Local customs and beliefs have a significant impact not only on the way that the disaster itself is perceived, but also on the credibility of different kinds of authority. Thus, on Montserrat, the role of religious leaders in urging the population to listen to the scientists was very significant (Donovan 2010).

There are also questions about the ways in which risk assessments are represented to policymakers and publics. Recent studies by Gigerenzer (Gigerenzer et al. 2005; Gigerenzer 2008) have shown that public and academic understanding of probabilities may be inadequate. Studies of this nature are needed in the geophysical hazards as part of the process of refining the expression of risk assessments. In particular, work with policymakers in this regard would be very significant. It is clear from recent experiences (notably Montserrat) that scientists' understanding of political systems and requirements is important in facilitating the flow of information in a crisis. The "ash crisis," for example, showed that scientists between the UK and Iceland could exchange information relatively easily, but beyond the scientific community, communication became more challenging. The communication process is particularly worthy of consideration in handling uncertainty, because it generates so much uncertainty itself through the misunderstandings of individuals.

One of the key challenges in promoting transboundary risk assessment and management is the coordination of scientific research and policy advice: enabling governments to locate appropriate experts rapidly, and providing experts with advice about their role (e.g., Renn 2008). This could be carried out by an international geophysical organization, perhaps following the paradigm of the World Meteorological Organization. Mapping available expertise, advising scientists on their role, translating research, and facilitating technological support for decision making would be useful functions that are currently lacking (e.g., Sparks 2007). Academic research in its traditional form is rarely accessible for those who might benefit from it. Conversely, scientists and social scientists may not be aware of their rights and

responsibilities when providing advice. There are also philosophical and epistemological considerations, since science can be misused as well as misinterpreted. An important step forward in the management of future geophysical disasters would seem to be an international body to maximize the usefulness of expertise and research for risk reduction.

## Conclusions

This chapter has documented the recent history of volcanic and seismic hazards and risk assessment. It has looked at some key case studies of earthquakes and eruptions that have had significant impacts on the science and social science of risk assessment and management. In particular, it has discussed the growth of uncertainty discourses and the relationship between science and policy in the volcanic and seismic contexts. This is a growing area of research that has great potential for further development in an interdisciplinary framework. Public and political perception of low-probability, high-impact risks is generally marred by preoccupation with matters of daily life (Gaillard 2008). This puts pressure on the research community to expand its efforts into the public sphere through education initiatives – and this is often a major part of the role of volcano observatories, for example. However, in many areas at risk from these hazards, the political channels for scientific advice simply do not exist. Preparedness is often linked to recent history of seismicity or volcanic activity. Given the catastrophic potential of geophysical hazards (e.g., Self 2006; Smil 2008), much further work is required uniting social and physical scientists with policymakers, NGOs, and publics. Communicating the nature of uncertainty is central in this process.

## References

- Aspinall WP et al (2002) The Montserrat volcano observatory: its evolution, organization, role and activities. *Geol Soc Lond Mem* 21(1):71–91
- Aspinall WP et al (2003) Evidence-based volcanology: application to eruption crises. *J Volcanol Geotherm Res* 128:273–285
- Aspinall WP (2006) Structured elicitation of expert judgement for probabilistic hazard and risk assessment in volcanic eruptions. In: Coles S, Mader HM, Connor C, Connor L (eds) *Statistics in volcanology*, vol 1, Special publications of Iavcei. Geological Society of London, London, pp 15–30
- Aspinall WP (2010) A route to more tractable expert advice. *Nature* 463:294–295
- Barberi F et al (2008) Volcanic risk perception in the Vesuvius population. *J Volcanol Geotherm Res* 172(3–4):244
- Barclay J et al (2008) Framing volcanic risk communication within disaster risk reduction: finding ways for the social and physical sciences to work together. *Geol Soc Lond Spec Publ* 305(1): 163–177
- Baxter PJ et al (2008) Emergency planning and mitigation at Vesuvius: a new evidence-based approach. *J Volcanol Geotherm Res* 178(3):454
- Beck U (1992) *Risk society: towards a new modernity*. Sage, New Delhi (German original, 1986)
- Beck U (2009) *World at risk*. Polity Press, Cambridge (German original 2007)
- Beck U (1999) *World risk society*. Polity Press, Cambridge
- Bird DK et al (2009) Public perception of jokulhaup hazard and risk in Iceland: implications for community education. *Int J Manag Decis Mak* 10(3–4):164–175
- Bird DK et al (2010) Volcanic risk and tourism in southern Iceland: implications for hazard, risk and emergency response education and training. *J Volcanol Geotherm Res* 189(1–2):33
- Bommer JJ, Abrahamson NA (2006) Why do modern probabilistic seismic hazard analyses often lead to increased hazard estimates? *Bull Seismol Soc Am* 96(6):1976–1977
- Castanos H, Lomnitz C (2002) PSHA: Is it science? *Eng Geol* 66(3–4):315–317

- Chester DK (2001) The 1755 Lisbon earthquake. *Prog Phys Geogr* 25(3):363–383
- Chester DK (2005) Theology and disaster studies: the need for dialogue. *J Volcanol Geotherm Res* 146(4):319–328
- Chester DK et al (2008) The importance of religion in shaping volcanic risk perception in Italy, with special reference to Vesuvius and Etna. *J Volcanol Geotherm Res* 172(3–4):216
- Cooke RM (1991) Experts in uncertainty: opinion and subjective probability in science. Oxford University Press, Oxford
- Corvalyn M (2008) Is probability the only coherent approach to uncertainty? *Risk Anal* 28(3):645–652
- Cox RT et al (2001) Neotectonics of the southeastern reelfoot rift zone margin, central United States, and implications for regional strain accommodation. *Geology* 29(5):419–422
- Cronin SJ et al (2004) Participatory methods of incorporating scientific with traditional knowledge for volcanic hazard management on Ambae Island, Vanuatu. *Bull Volcanol* 66(7):652
- De Finetti B (1974) Theory of probability. Wiley, London
- Donovan AR (2010) Emerald and andesite: volcanology at the policy interface on Montserrat. Unpublished PhD thesis, University of Cambridge, Cambridge
- Donovan AR, Oppenheimer C, Bravo M (2011) Rationalising a crisis through literature: Montserratian verse and the descriptive reconstruction of an island. *J Volcanol Geotherm Res* 203(3–4):87–101
- Donovan AR, Oppenheimer C (2010) The 2010 Eyjafjallajökull eruption and the reconstruction of Geography. *The Geogr J* 177(1):4–11
- Dove MR (2008) Perception of volcanic eruption as agent of change on Merapi volcano, central Java. *J Volcanol Geotherm Res* 172(3–4):329
- Druitt TH, Kokelaar BP (eds) (2002) The eruption of the Soufrière hills volcano, Montserrat, from 1995 to 1999, vol 21. Geological Society of London, London
- Erfurt-Cooper P, Cooper M (2010) Volcano and geothermal tourism: sustainable geo-resources for leisure and recreation. Earthscan, London
- Esteve L (1969) Seismicity prediction: a Bayesian approach. In: Proceedings of the fourth world conference on earthquake engineering, Santiago
- Ewert JW, Newhall CG (2004) Status and challenges of volcano monitoring worldwide. In: Proceedings of the 2nd international conference on volcanic ash and aviation safety, 21–24 June, 2004. Office of the Federal Coordinator for Meteorological Services and Supporting Research, Alexandria, session 2, p 9–14
- Faenza L et al (2010) Bayesian inference on earthquake size distribution: a case study in Italy. *Bull Seismol Soc Am* 100(1):349–363
- Finch RH (1943) The seismic prelude to the 1942 eruption of Mauna Loa. *Bull Seismol Soc Am* 33(4):237–241
- Fischer F (2004) Are scientists irrational? Risk assessment in practical reason. In: Scoones I, Leach M, Wynne B (eds) Science and citizens: globalisation and the challenge of engagement. Zed Books, London, pp 54–65
- Fiske R (1984) Volcanologists, journalists and the concerned local public: a tale of two crises in the eastern Caribbean. In: Explosive volcanism: inception, evolution and hazards. National Academy Press, Washington, DC
- Francis P, Oppenheimer C (2004) Volcanoes. Oxford University Press, Oxford
- Gaillard J-C (2008) Alternative paradigms of volcanic risk perception: the case of Mt. Pinatubo in the Philippines. *J Volcanol Geotherm Res* 172(3–4):315
- Geller RJ (1997) Earthquake prediction: a critical review. *Geophys J Int* 131(3):425–450
- Gelman A (2008) Objections to Bayesian statistics. *Bayesian Analysis* 3(3):445–450
- Gelman A et al (2004) Bayesian data analysis. Chapman and Hall/CRC Press, Boca Raton
- Gibbons M et al (1994) The New production of knowledge: the dynamics of science and research in contemporary societies. Sage, London
- Gigerenzer G (2008) Why heuristics work. *Perspect Psychol Sci* 3(1):20
- Gigerenzer G et al (2005) ‘A 30% Chance of rain tomorrow’: how does the public understand probabilistic weather forecasts? *Risk Anal* 25(3): 623–629
- Hamilton W (1772) Observations of mount Vesuvius mount Etna and other volcanoes. T. Cradell, London
- Haynes K et al (2007) The issue of trust and its influence on risk communication during a volcanic crisis. *Bull Volcanol* 70(5):605–621
- Haynes K et al (2008) Whose reality counts? Factors affecting the perception of volcanic risk. *J Volcanol Geotherm Res* 172(3–4):259
- Hillerbrand R, Ghil M (2008) Anthropogenic climate change: scientific uncertainties and moral dilemmas. *Physica D* 237:2132–2138
- Hincks T (2005) Probabilistic volcano hazard and risk assessment. PhD dissertation, University of Bristol, Bristol
- Hulme M (2009) Why we disagree about climate change: understanding controversy, inaction and opportunity. Cambridge University Press, Cambridge
- Hulme M, Mahony M (2010) Climate change: what do we know about the Ipcc? *Prog Phys Geogr* 34(5): 705–718
- Huppert HE, Sparks RSJ (2006) Extreme natural hazards: population growth, globalization and

- environmental change. *Phil Trans R Soc A Math Phys Eng Sci* 364(1845):1875–1888
- Ives RL (1937) Volcanic eruptions predicted. *Sci News Lett* 32(860):218
- Jackson J (2006) Fatal attraction: living with earthquakes, the growth of villages into megacities, and earthquake vulnerability in the modern world. *Phil Trans R Soc A: Math Phys Eng Sci* 364(1845):1911–1925
- Johnston AC, Nava SJ (1985) Recurrence rates and probability estimates for the New Madrid seismic zone. *J Geophys Res* 90(B8):6737
- Kahan D (2010) Fixing the communications failure. *Nature* 463(7279):296
- Kahan DM et al (2009) Cultural cognition of the risks and benefits of nanotechnology. *Nat Nano* 4(2):87
- Kanamori H (2006) Lessons from the 2004 Sumatra-Andaman earthquake. *Phil Trans R Soc A: Math Phys Eng Sci* 364(1845):1927–1945
- Kilburn CRJ (2003) Multiscale fracturing as a key to forecasting volcanic eruptions. *J Volcanol Geotherm Res* 125(3–4):271
- Kilburn CRJ, Sammonds PR (2005) Maximum warning times for imminent volcanic eruptions. *Geophys Res Lett* 32(24):L24313
- Krinitzsky EL (1995) Deterministic versus probabilistic seismic hazard assessment for critical structures. *Eng Geol* 40(1–2):1–7
- Lacroix A (1904) La Montagne Pele E Ses Eruptions. Masson, Paris
- Leonard GS et al (2008) Developing effective warning systems: ongoing research at Ruapehu volcano, New Zealand. *J Volcanol Geotherm Res* 172(3–4):199
- Lindell MK, Perry RW (1993) Risk area residents' changing perceptions of volcano hazard at Mt St Helens. In: Newec JJ et al (eds) Prediction and perception of natural hazards. Springer, Berlin, pp 159–166
- Lowenstern JB et al (2006) Monitoring super-volcanoes: geophysical and geochemical signals at yellowstone and other large caldera systems. *Phil Trans R Soc A: Math Phys Eng Sci* 364(1845):2055–2072
- MacGregor A (1949) Prediction in relation to seismo-volcanic phenomena in the Caribbean volcanic Arc. *Bull Volcanol* 8(1):69
- Marzocchi W et al (2006) A quantitative model for volcanic hazard assessment. In: Coles S, Mader HM, Connor C, Connor L (eds) Statistics in volcanology, vol 1, Special publications of Iavcei. Geological Society, London
- Marzocchi W et al (2007) BET\_EF: a probabilistic tool for long- and short-term eruption forecasting. *Bull Volcanol* 70(5):623–632
- Marzocchi W, Woo G (2009) Principles of volcanic risk metrics: theory and the case study of mount Vesuvius and Campi Flegrei, Italy. *J Geophys Res* 114(B3):B03213
- Mason BM et al (2004) The size and frequency of the largest explosive eruptions on Earth. *Bull Volcanol* 66(8):735–748
- Matthews RAJ (1997) Decision-theoretic limits on earthquake prediction. *Geophys J Int* 131(3):526–529
- McGuire RK (2001) Deterministic vs probabilistic seismic hazards and risk. *Soil Dyn Earthq Eng* 21(5):377–384
- McGuire RK (2008) Probabilistic seismic hazard analysis: early history. *Earthq Eng Struct Dyn* 37:329–338
- McKenzie D, Parker RL (1967) The north pacific: an example of tectonics on a sphere. *Nature* 216:1276–1280
- Melnik O, Sparks R (1999) Nonlinear dynamics of lava dome extrusion. *Nature* 402:37–41
- Mitchell T (2006) Building a disaster resilient future: lessons from participatory research on St Kitts and Montserrat. Unpublished PhD thesis, University College London, London
- Moore JG, Rice CJ (1984) Chronology and character of the May 18, 1980, explosive eruptions of mount St Helens. In: Explosive volcanism: inception, evolution and hazards. National Academy Press, Washington, DC
- Morgan WJ (1968) Rises, trenches, great faults and crustal blocks. *J Geophys Res* 73(6):1959–1982
- Neri A et al (2008) Developing an event tree for probabilistic hazard and risk assessment at Vesuvius. *J Volcanol Geotherm Res* 178(3):397
- Newhall C, Hoblitt RP (2002) Constructing event trees for volcanic crises. *Bull Volcanol* 64:3–20
- Newhall C, Punongbayan R (1996) The narrow margin of successful volcanic-risk mitigation. In: Tilling RI, Scarpa R (eds) Monitoring and mitigation of volcano hazards. Springer, New York, pp 807–832
- Newhall CG et al (1999) Professional conduct of scientists during volcanic crises. *Bull Volcanol* 60:323–334
- O'Hagan A et al (2006) Uncertain judgements: eliciting Experts' probabilities. Wiley, London
- Oreskes N et al (1994) Verification, validation, and confirmation of numerical models in the earth sciences. *Science* 263(5147):641–646
- Oreskes N, Le Grand H (eds) (2003) Plate tectonics: an insider's history of the modern theory of the earth. Westview Press, Oxford
- Pattullo P (2000) Fire from the mountain: the tragedy of Montserrat and the betrayal of its people. Constable, London
- Perry RW, Greene M (1983) Citizen response to volcanic eruptions: the case of Mount St Helens. Irvington Press, New York
- Perry RW, Lindell MK (2008) Volcanic risk perception and adjustment in a multi-hazard environment. *J Volcanol Geotherm Res* 172(3–4):170

- Pidgeon N et al (2003) The social amplification of risk. Cambridge University Press, Cambridge
- Reasenberg PA, Simpson RW (1992) Response of regional seismicity to the static stress change produced by the Loma Prieta earthquake. *Science* 255(5052): 1687–1690
- Renn O (2008) Risk governance. Earthscan, London
- Roeser S (2007) Ethical intuitions about risk. *Saf Sci* 11(3):1–13
- Roeser S (2010) Intuitions, emotions and Gut reactions in decisions about risks: towards a different interpretation of “neuroethics”. *J Risk Res* 13(2): 175–190
- Roeser S (2009) The relation between cognition and affect in moral judgements about risks. In: Asveld L, Roeser S (eds) The ethics of technological risk. Earthscan, London
- Roeser S (2006) The role of emotions in judging the moral acceptability of risks. *Saf Sci* 44(8):689
- Sackett DL, Rosenberg WMC (1995) On the need for evidence-based medicine. *J Public Health* 17(3): 330–334
- Sackett DL et al (1996) Evidence based medicine: what it is and what it isn't. *BMJ* 312(7023):71–72
- Self S (2006) The effects and consequences of very large explosive volcanic eruptions. *Phil Trans R Soc A: Math Phys Eng Sci* 364(1845):2073–2097
- Shackley S, Wynne B (1996) Representing uncertainty in global climate change science and policy: boundary-ordering devices and authority. *Sci Technol Hum Value* 21(3):275–302
- Shackley S et al (1998) Uncertainty, complexity and concepts of good science in climate change modeling: are GCMs the best tools? *Clim Chang* 38(2):159
- Sieh K (2006) Sumatran megathrust earthquakes: from science to saving lives. *Phil Trans R Soc A: Math Phys Eng Sci* 364(1845):1947–1963
- Sigurdsson H (1999) Melting the earth: the history of ideas on volcanic eruptions. Oxford University Press, Oxford
- Sjoberg L (2008) Antagonism, trust and perceived risk. *Risk Manage* 10:32–55
- Sjoberg L (2000) Factors in risk perception. *Risk Anal* 20:1–11
- Slovic P (2000) The perception of risk. Earthscan, London
- Slovic P et al (2004) Risk as analysis and risk as feelings: some thoughts about affect, reason, risk, and rationality. *Risk Anal* 24(2):311
- Smil V (2008) Global catastrophes and trends: the next fifty years. MIT Press, Cambridge
- Solana MC et al (2008) Communicating eruption and hazard forecasts on Vesuvius, southern Italy. *J Volcanol Geotherm Res* 172(3–4):308
- Sparks RSJ (2003) Forecasting volcanic eruptions. *Earth Planet Sc Lett* 210:1–15
- Sparks RSJ (2007) Use the calm between the storms. *Nature* 450:354
- Spence R et al (2005) Residential building and occupant vulnerability to tephra fall. *Nat Hazards Earth Syst Sci* 5:477–495
- Spence R et al (2008) Modelling the impact of a hypothetical Sub-plinian eruption at La soufrière of Guadeloupe (lesser Antilles). *J Volcanol Geotherm Res* 178(3):516
- Spence R (2009) Earthquake risk mitigation: the global challenge. In: Tankut AT (ed) Earthquakes and tsunamis, vol 11, Geotechnical, geological and earthquake engineering. Springer, Dordrecht
- Spiegelhalter D, Riesch H (in review) Don't know, can't know: embracing scientific uncertainty when analyzing risks
- Stein RS et al (2006) A New probabilistic seismic hazard assessment for greater Tokyo. *Phil Trans R Soc A: Math Phys Eng Sci* 364(1845):1965–1988
- Stirling A (2007) Risk, precaution and science: towards a more constructive policy debate. *EMBO Rep* 8(4):309–315
- Stirling A (2008) Opening up and closing down. *Sci Technol Hum Value* 33(2):262–294
- Taleb NN (2007) The black swan: the impact of the highly improbable. Allen Lane, London
- Tayag JC, Punongbayan RS (1994) Volcanic disaster mitigation in the Philippines: experience from mount Pinatubo. *Disasters* 18(1):1–15
- Tazieff H (1977) La Soufrière, volcanology and forecasting. *Nature* 269:96–97
- Tilling RI (2008) The critical role of volcano monitoring in risk reduction. *Adv Geosci* 14:3–11
- Voight B (1988) A method for prediction of volcanic eruptions. *Nature* 332:125–130
- Wadge G, Isaacs M (1988) Mapping the volcanic hazards from the Soufrière hills volcano, Montserrat, west indies, using an image processor. *J Geol Soc Lond* 145:541–551
- Waesche HH (1942) Ground tilt at Kilauea volcano. *J Geol* 50(6):643
- Wang Z (2009) Seismic hazard vs seismic risk. *Seismol Res Lett* 80(5):673–674
- Wang Z (2010) Seismic hazard assessment: issues and alternatives. *Pure Appl Geophys* 168(1–2):11–25. doi:[10.1007/s00024-010-0148-3](https://doi.org/10.1007/s00024-010-0148-3)
- Wilson T et al (2009) Vulnerability of farm water supply systems to volcanic Ash fall. *Environ Earth Sci* 61(4):675
- Winchester S (2003) Krakatoa: the day the world exploded. Viking, Camberwell
- Wisner B et al (2004) At risk: natural hazards, people's vulnerability and disasters. Routledge, London

- Witham CS, Oppenheimer C (2004) Mortality in England during the 1783-4 Laki Craters eruption. *Bull Volcanol* 67(1):15–26
- Woo G (1999) The mathematics of natural catastrophes. Imperial College Press, London
- Wynne B (1992) Uncertainty and environmental learning: reconceiving science and policy in the preventive paradigm. *Glob Environ Chang* 2(2): 111–127
- Young DA (2003) Mind over magma: the story of igneous petrology. Princeton University Press, Princeton
- Zerefos CS et al (2007) Atmospheric effects of volcanic eruptions as seen by famous artists and depicted in their paintings. *Atmos Chem Phys Discuss* 7(2):5145–5172
- Zuccaro G et al (2008) Impact of explosive eruption scenarios at Vesuvius. *J Volcanol Geotherm Res* 178(3):416



# Part 3

---

## **Decision Theory and Risk**



# 15 A Rational Approach to Risk? Bayesian Decision Theory

Claus Beisbart

Technische Universität Dortmund, Dortmund, Germany

<i>Introduction</i> .....	376
<i>History</i> .....	382
The Von Neumann–Morgenstern Representation Theorem .....	382
Ramsey’s Approach .....	387
Savage’s Theory .....	391
The Bolker–Jeffrey Theory .....	394
<i>Current Research: Bayesian Decision Theory and Risk</i> .....	395
<i>Further Research</i> .....	398

**Abstract:** The aim of this chapter is to provide an overview over decision theory, in particular Bayesian decision theory, to explain its main ideas with some emphasis on risk and to flag the most important controversies about the theory. The paper starts out with the St. Petersburg paradox and motivates the idea that rational agents maximize expected utility. Since “rationality” is ambiguous, the attention is restricted to a core notion of rationality that evaluates an agent’s choices in relation to her desires and beliefs. Accordingly, in Bayesian decision theory, the expected utility arises from utilities and probabilities that measure the strengths of the agent’s desires and beliefs, respectively. This raises two questions: (1) Why do rational agents maximize expected utility? (2) How can the strengths of the agent’s desires and beliefs be measured? Both questions are commonly answered in terms of representation theorems. Such theorems show that the strengths of the agent’s beliefs and desires can be represented in terms of numerical probabilities and utilities if the choices she would take in hypothetical cases fulfill certain conditions. The theorems further entail that an option is preferred to another if and only if it has a higher expected utility than the latter. This chapter covers results by von Neumann–Morgenstern, Ramsey, Savage, and Bolker–Jeffrey. It discusses the consequences for choices that are risky in a nontechnical sense and concludes by pointing out major controversies concerning Bayesian decision theory.

## Introduction

---

If a person is rational, she will choose an option that maximizes what she expects to gain from her choice. This is, in very rough terms, the core of much decision theory, particularly of Bayesian decision theory. In more mathematical terms, it is claimed that a rational agent maximizes her expected utility. The expected utility arises from utilities that express the extent to which various outcomes are desired by the agent and from probabilities. Proponents of Bayesian decision theory – Bayesians, for short – stress that the probabilities express the degrees to which the agent believes certain propositions.

Decision theory is extremely influential in philosophy, economics, in the foundations of statistics (cf. French and Ríos Insua 2000; Robert 2007), and in other fields. It also shapes much thinking about risk, not just because risk is a technical term in decision theory, but also because it suggests advice how we should rationally choose from options that are risky in a nontechnical sense. The aims of this chapter are to unfold the idea that rational agents maximize expected utility, to review the most important results and controversies about this idea, and to discuss its consequences for risk. The emphasis is on Bayesian decision theory, but I will also cover other approaches from what I call *standard decision theory*. This term is my umbrella term for theories that take rational agents to maximize expected utility (see Sugden 2004 for alternatives).

Decision theory is centered on mathematical results and its achievements cannot be appreciated without some mathematics. But the mathematics can get quite involved and cannot be outlined in an chapter like this. I will therefore take a middle course and informally describe the most important mathematical results using examples. My aim is to get across the basic ideas rather than to make mathematically precise statements. I recommend Raiffa (1968), Fishburn (1979), Kreps (1988), Hammond (1998a), Hammond (1998b) and Sugden (2004) for more formal reviews of decision theory; see also Barberà et al (1998, 2004) and Peterson (2009).

The plan of this chapter is as follows: The remainder of the introduction motivates the idea that expected utility is maximized by rational agents. The section about history explains the most

famous related approaches in decision theory. I discuss the implications about risk in the section about current research. A survey of philosophical issues about decision theory is provided in the last section.

To begin, suppose that Peter wants to sell his old car. He can either sell it to a lawyer at a price of \$3,000. There is a chance of 20% that the car will break down in the next year, and to sell the car to the lawyer, Peter has to agree to pay him back \$2,000 in case the car breaks down. Alternatively, Peter can sell the car to his old friend Mary at a price of \$2,000. If she gets the car at this price, she agrees not to claim any money back if the car breaks down. What should Peter do? How should he decide?

One way to give Peter is advice is to consider the money that Peter will gain. Let us do so using a table. Peter has two *options* – selling the car to the lawyer and selling it to Mary. Depending on the *conditions* – whether the car will break down in the next year or not – the options lead to different *outcomes*. In  *Table 15.1*, each row represents an option and each column represents a condition. Accordingly, each box stands for an outcome that arises by combining an option and a condition. In each box, we note the money that Peter gains given an outcome. In the lower half of each box, we additionally put down the probability that the condition corresponding to the outcome obtains. For instance, the upper left box represents the outcome that arises if Peter sells the car to the lawyer and if the car does not break down. Peter will then obtain \$3,000 all in all. The probability of the car not breaking down is .8. Note that the probabilities in each row sum up to 1 – the conditions jointly exhaust the space of possibilities.

In the example, we assume that the probability of the car breaking down is .2, independently of whether the car is sold to the lawyer or to Mary. Accordingly, each box from the same column has the same probability. But the probability that the condition obtains may also depend on the option that Peter chooses. Maybe the car is more likely to break down if it is sold to the lawyer, because he uses the car more often than Mary would do. In this case, the probability in each box would have to be the conditional probability of the condition, given that Peter takes the option corresponding to the box.

Good advice for Peter could focus on the *expected monetary value* of each option. Consider selling the car to the lawyer first. With a chance of 80%, the car will not break down in the next year and Peter will not pay back anything, in which case the monetary value of the option for Peter is \$3,000. With a 20% chance, the car will break down, in which case the monetary value of the option for Peter is  $\$3,000 - \$2,000 = \$1,000$ . Thus, on average, the monetary value of the option for Peter is  $\$3,000 \times .8 + \$1,000 \times .2 = \$2,600$ . This is the expected monetary value of the option. If Peter were faced with the same decision many times and would always take the first option, he would very likely gain \$2,600 on average. Consider now the other option.

 **Table 15.1**  
A simple decision problem

	Car does not break down	Car breaks down
Sell car to lawyer	\$3,000 .8	\$1,000 .2
Sell car to Mary	\$2,000 .8	\$2,000 .2

The expected monetary value of the option is \$2,000 independently on whether the car breaks down or not. Thus, the expected monetary value is \$2,000. This is less than the expected monetary value of the first option. Peter is thus better off if he chooses the first option. It thus seems that Peter should take the first option.

When we generalize this idea, we come up with the following *decision rule*, call it the *maximum expected monetary value-rule* (MEMV):

*MEMV.* An agent should choose the option with the maximum expected monetary value. If several options maximize the expected monetary value, she is permitted to take any of these options.

But is MEMV a sensible rule for making decisions quite generally? Obviously, it is not. One reason is that money is not everything that is important about the options. There may be additional benefits for Peter if he sells his car to his old friend Mary – she will invite him to her famous parties, say. Now we may perhaps take this additional benefit into account by translating it into money too. But this would not get us very far for other reasons. The point is simply that money does not really count what matters to us. Suppose, for instance, that Peter needs additional \$2,000 at the end of next year and that he is in deep trouble if he does not have the money at that time. In this case, Peter should probably choose the second option because it gives him \$2,000 for sure. If he chooses the other option, on the contrary, there is some chance that he does not have the additional \$2,000 at the end of the year and thus runs into deep trouble (Resnik 1987, pp. 85–88; cf. also Lumer 2002, pp. 2–4).

That maximizing the expected monetary value is not a sensible idea can also be made vivid using the so-called *St. Petersburg paradox*. This point was made in a paper that was written by Daniel Bernoulli in 1738 and that is often thought to be the first work in utility theory (Bernoulli 1738/1954, crucial ideas in the paper are due to Gabriel Cramer). Suppose, you can buy a ticket for the St. Petersburg gamble, which is defined as follows. A fair coin is flipped until it lands heads for the first time. If  $n$  is the number of flips needed to produce the first heads, you are given  $\$2^n$ . Thus, you obtain \$2 in case the coin lands heads at the first flip; you obtain \$4 if the coin first lands tails and then heads, and so on. As the coin is fair, the probability to get \$2 is 1/2, the probability to get \$4 is 1/4, and so on. The expected monetary value of the gamble is thus

$$(\$2/2 + 2^2/2^2 + 2^3/2^3 + \dots),$$

which is infinity. The ticket for the gamble has a finite price. You have to decide between buying the ticket or not buying the ticket. If you maximize the expected monetary value, you should buy the ticket whatever price it has. The reason is that the expected monetary value of buying the ticket is infinite; the gamble will on average give you infinitely many \$, which will compensate for the price of the ticket. On the other hand, the expected monetary value of not buying the ticket is zero. The paradox arises because it seems insane to pay much money for a ticket intuitively. It can be shown, for instance, that you will obtain \$4 or less with a probability of 75%.

Bernoulli resolved the paradox by suggesting that the desirability of some amount of money does not increase in the same way as the amount itself increases. Put differently, the value of additional \$1,000 for you depends on the amount of money that you already own. \$1,000 is very valuable to you if you do not have any other money, but it does not make much of a difference if you have already \$1,000,000. This is captured by the *Law of Diminishing Marginal Utility*, which is well-known from economics (e.g., Samuelson and Nordhaus 2001, p. 86).

Bernoulli suggested that the desirability that some amount of money has for you is proportional not to this amount, but rather to its logarithm. That is, the desirability of \$1,000,000 is not just 1,000 times the desirability of \$1,000, but only twice the latter. One can show that, under this assumption, the expected monetary value of the St. Petersburg gamble is finite and that the gamble is not worth any price. The counterintuitive consequences of the St. Petersburg paradox are avoided. To avoid similar counterintuitive consequences concerning gambles that are similar to the St. Petersburg gamble, the desirability of money has to be bounded from above (Menger 1931; see also Resnik 1987, pp. 107–109).

We conclude that MEMV is not a good idea. But it seems easy to amend MEMV. We require that not the expected monetary value, but rather the expected *desirability* of a choice be maximized. Let us assume that the desirability of an outcome can be measured in quantitative terms using *utilities* (more on this shortly). Suppose further that the utilities of our decision problem are as in  [Table 15.2](#). Note that the utility of an outcome is not proportional to the money it affords Peter.

To give advice to Peter, let us calculate the *expected utility* (rather than the expected monetary value) for each option. We obtain an expected utility of  $.8 \times 200 + .2 \times 100 = 180$  for the first option and an expected utility of 190 for the second option. Thus, Peter should sell the car to Mary.

We have now applied the *maximum expected utility-rule (MEU)*:

*MEU*. An agent should choose the option with the maximum expected utility. If several options maximize the expected utility, she is permitted to take any of these options.

Utility maximization avoids some pitfalls of MEMV. But the question is nevertheless

(Q) Is MEU is defensible? Should people really maximize expected utility?

Alternative rules are possible. For instance, the agent could focus on the minimum utility possible under an option and then try to maximize this minimum amongst the options. This is the famous Maximin rule (Resnik 1987, pp. 26–28, see also Rawls 1971, Part I, particularly §26 and Gardiner 2006).

The question whether MEU is defensible (Q) turns on two other questions.

(Qa) Where do the utilities come from? I have said that a utility measures the desirability of an outcome, but what kind of desirability are we talking of? Are we concerned with moral desirability, with desirability for Peter – call this prudential desirability – or with some other desirability? And how is the relevant desirability measured using numbers?

(Qb) Where do the probabilities come from? How do we obtain their values?

 **Table 15.2**  
The decision problem

	Car does not break down	Car breaks down
Sell car to lawyer	200	100
	.8	.2
Sell car to Mary	190	190
	.8	.2

The numbers in the upper halves of the boxes represent the utilities of the outcomes

Clearly, whether MEU is defensible turns on what the utilities and probabilities express and how their values are determined.

We cannot answer the questions (Q), (Qa), and (Qb) unless we decide what the focus our theory should be and what kind of advice we are interested in. Let us first think about *desirability* (Qa). We may decide to understand “desirability” as *moral* desirability. Desirability would then have to be the kind of value that matters for morality, and the utilities would measure it. With such a notion of desirability, MEU would become a moral principle and its “should” would concern the moral rightness of choices. As is well-known, utilitarianism defends something like MEU as the principle of morality (for utilitarianism see e.g., Sen and Williams 1982; Glover 1990; Shaw 1999 and Darwall 2002). If we understand MEU as a candidate for a moral principle, the decisive questions are whether we can quantify moral desirability using utilities (Qa) and whether maximizing moral desirability gives a defensible account of moral rightness (Q).

Alternatively, we may want to develop a theory of *prudential* reason. The utilities would then have to measure how much the outcomes contribute to the well-being of some fixed person (presumably the agent). The decisive questions then are whether the well-being of a person can be measured using utilities (Qa; see Griffin 1989, Part II) and whether maximizing one’s well-being in this sense is a defensible account of prudential reason (Q).

Decision theorists take yet another line. They think of desirability in terms of what the agent desires. An outcome is taken the more desirable, the more it is desired. Decision theorists thus use MEU to judge the agent’s choices by her own lights, i.e., from the standpoint of her own desires. A related judgment concerns the rationality of a choice – a *rational* agent will choose that option that maximally reflects what she wants. Thus, MEU is about what agents should do rationally. The decisive questions then are whether the strengths of a person’s desires can be measured using utilities (Qa) and whether maximizing desirability in this sense is a defensible account of rational choice (Q).

The term “rational” is of course ambiguous. Decision theorists conceive of rationality in a fairly minimal sense. This sense of rationality may be called *means-ends rationality* or *instrumental rationality* (recent philosophical work about means-ends rationality includes Bratman 1987; Kordsgaard 1997; Wallace 2001; Broome 2002; Smith 2004 and Raz 2005). Very roughly, an agent is means-ends rational if and only if (iff, for short) she takes means that she regards to be appropriate to realize her ends or to satisfy her desires. Conversely, we can criticize an agent as not fully means-ends rational if, in a certain situation, only one of her ends is relevant and if she does not take an option that she takes to be necessary and sufficient to realize it. This example is often used to explain instrumental rationality, but the assumption that there is only one relevant aim and that exactly one option is taken to be necessary and sufficient to realize the aim is completely unrealistic in most cases. Typically, there are many things that the agent wants, and each option will make a difference to many things that the agent wants. What then does means-ends rationality require in such a case? MEU is a candidate for spelling out means-ends rationality quite generally. The idea is that a rational agent maximizes the expectation value of what she wants.

Read in this way, MEU is still normative. It is about what one should do in some sense of “should.” There is an alternative way to understand MEU. The idea is that MEU constrains the choices of agents that are rational in a technical sense. I will come back to a related reading of MEU in the last section of this chapter.

Let us now think about the *probabilities* and where they come from (Qb). A theorist faces two basic options. One can either let the probabilities reflect real-world chances. This is particularly

plausible for gambles such as the St. Petersburg gamble. Presumably, each coin has a well-defined objective probability to land heads. The alternative is to say that the probabilities reflect the agent's own expectations. The probabilities will then express how likely the agent takes it to be that an option leads to a specific outcome. In slightly different terms, they measure the extent to which the agent believes that an option leads to a specific outcome. They measure degrees of belief or levels of confidence (see Kyburg and Smokler 1964 for an anthology).

*Bayesian* decision theory opts for the second alternative. That is, Bayesian decision theorists use MEU to judge an agent by her own lights not just as far as the utilities are concerned, but also concerning the probabilities. The choices of the agent are judged in relation to her own desires and beliefs. It is arguable that we are now talking about the centerpiece of rational choice and rational agency. The reason is that we only call a choice irrational if the agent fails by her own lights, i.e., relative to her own desires and beliefs. On the contrary, if an agent does what she *falsely takes* to serve her ends maximally, her choice is not irrational, but rather ineffective (cf. Scanlon 1998, Chap. 1).

Here then is the core of Bayesian decision theory (BDT):

*BDT.* An option is rational to choose iff it maximizes the expected utility of the agent (if several options tie at the maximum expected utility, each of them is rational to choose). The expected utility of an option is the expectation value of the utility of the outcome, given this option. Here the utility of an outcome measures the degree to which the outcome is desired by the agent. The expectation value is formed using probabilities that reflect the degrees of confidence that the agent assigns to an outcome, given that a particular option is chosen.

Three warnings are in order. First, no presumption is made concerning the content that the desires of the agent have. It is neither assumed that the agent only cares about her own good nor that her desires are only directed at her own pleasure. This may or may not be so. As has often been pointed out, it is unfortunate to speak of utilities in decision theory given that hedonic accounts of utility dominated in early utilitarianism.

Second, to form an expectation value of a utility, one has to have probabilities. What is most important for BDT from a conceptual point of view is that the probabilities reflect the agent's own beliefs about the world – otherwise we would not be talking about rationality in the core sense picked by Bayesians. As already indicated, Bayesians take probabilities to measure *degrees of belief*. The basic assumption is that belief comes in degrees. This can be motivated using the preface paradox (Makinson 1965; Foley 1992; Hawthorne and Bovens 1999). The agent's probability for a proposition is then taken to be the degree to which the agent believes the proposition. A proposition that is endorsed as true by the agent has a degree of belief or probability 1. A proposition that is believed to be false or disbelieved has 0 as its degree of belief or probability. Propositions that the agent is not certain about have a degree of belief strictly between 0 and 1 (it is not clear whether there is also a type of belief that does not come in degrees and how it is related to credences that do have degrees). The idea that probabilities measure degrees of belief may also be employed more generally to explain what probabilities are and what probabilistic statements mean. Views that take probabilistic statements to express degrees of belief of the speaker are called subjectivist. The important point to note is that BDT does not entail a subjectivist view for every probability. It is compatible with BDT that some probabilities are objective, for instance, that they reflect propensities (dispositions) of real-world systems. BDT is not meant to yield a general interpretation of probabilities

(for interpretations of probabilities see Fine 1973, Howson 1995, Gillies 2000; Hacking 2001; Mellor 2005 and Hájek 2010; see also Hacking 1975, 1990; Skyrms 1999; Jeffrey 2004 and Howson and Urbach 2006).

Third, BDT has an “if,” and not an “iff.” It claims a necessary condition of rational choice, but we need not think that this is the only constraint of rationality. In the example, Peter may not be fully rational if another possible constraint is violated, e.g., if his utilities would not stand certain reflection. Maybe, the utilities would change and become more rational if Peter tried to imagine what it means to sell the car to Mary. Additionally, Peter may not be rational if his probabilities do not properly reflect his evidence (more on this later). Whether there are such additional constraints of rationality is not at issue when we think about BDT. In particular, the central results of standard decision theory apply to one single choice. They do not constrain how utilities and probabilities (or desires and beliefs) should rationally change (see below for qualifications). We may thus say that BDT is concerned with synchronic aspects of rationality. One reason is that decision theory and BDT in particular have chosen to focus on a narrow core of practical rationality to begin with.

BDT is a development of MEU. We have obtained BDT by determining the focus of the theory. This is a first step to answer our above questions. But we have not yet answered these questions. So is BDT a defensible claim about rationality (Q)? And how can one measure the extent to which something is desired using utilities (Qa)? Finally, how can one measure degrees of belief using probabilities (Qb)?

To answer these questions is the point of much formal decision theory. An answer is often provided in terms of a *representation theorem*. Most representation theorems come in two parts. First, there is an existence claim according to which one can represent the strengths of desires and beliefs using numbers, viz. utilities and probabilities. The representation is such that the rational option to choose maximizes expected utility. Second, there is a uniqueness claim according to which the representation is unique in some sense. Of course, a unique representation is only possible under certain assumptions. These assumptions are a matter of much philosophical discussion.

In the following section, I will present four approaches to obtain a representation theorem. As we will see, some of the approaches do not really get us as far as BDT. I will nevertheless survey them because they are important in decision theory.

## History

---

### The Von Neumann–Morgenstern Representation Theorem

---

The most important step in utility theory after Bernoulli's 1738 paper was presumably the representation theorem by John von Neumann and Oskar Morgenstern (1947) (cf. Fishburn 1989). Their aim was to provide foundations for utility theory, which was at that time already used in economics. It was notoriously unclear though what utilities are. The early hedonic conceptions due to Bentham (1789) and Mill (1863) are problematic for both descriptive and normative purposes. The crucial idea of von Neumann and Morgenstern is to identify the utility of an outcome with the degree to which it is desired. They provide a precise method to obtain numerical utilities. Their theorem does not provide a method to measure degrees of belief though because the probabilities are assumed as given. For similar results see Marschak (1950), Herstein and

Milnor (1953), Hausner (1954) and Luce and Raiffa (1957, Chap. 2). In the following, I will not follow von Neumann and Morgenstern, but rather present a more elegant version of a von Neumann–Morgenstern-style representation theorem due to Jensen (1967) and rely on Hammond (1998a, pp. 152–164).

Let us start from the idea that rational agents maximize expected utility. To calculate the expected utility of an option, we need a utility for each possible outcome. In mathematical terms, each outcome  $a$  has to be mapped to a real number  $u(a)$ , which is called its utility.  $u(a)$  is supposed to *represent* the extent to which the outcome  $a$  is desired. Some outcomes are more strongly desired than others or (*strictly*) *preferred* to them, as I shall also say. Let us introduce some abbreviations and let “ $a \succ b$ ” denote that  $a$  is preferred to  $b$ . That  $a$  is preferred to  $b$  means that the agent will always choose  $a$  if presented with a choice between outcomes  $a$  and  $b$ . Preferences and thus desires are thus connected to hypothetical choices (see Fehige and Wessels 1998 for a collection about preferences).

If the utilities are to represent the extents to which outcomes are desired by the agent, or her preferences, for short, the utility of an outcome  $a$ ,  $u(a)$ , should be larger than that of an outcome  $b$ ,  $u(b)$ , if  $a$  is preferred to  $b$  (i.e., if  $a \succ b$ ). Let us therefore require that

$$u(a) > u(b) \text{ iff } a \succ b. \quad (15.1)$$

Here, the first “ $>$ ” sign denotes the “larger than”-relation among numbers.

As is well-known, one cannot always find a mapping  $u$  that satisfies our requirement. Suppose, for instance, that the agent has cyclic preferences. She prefers  $a$  to  $b$ ,  $b$  to  $c$ , and  $c$  to  $a$ . Any mapping from outcomes into real numbers that respects our requirement of representation would have to obey:

$$u(a) > u(b), \quad u(b) > u(c), \quad u(c) > u(a).$$

But this is impossible for any numbers  $u(a)$ ,  $u(b)$  and  $u(c)$  since it implies  $u(a) > u(a)$ , which is plainly false. The agent must not have any cyclic preferences if a representation is to be possible.

Now to have cyclic preferences is in fact odd. Suppose, Peter prefers a salad to a pizza, a pizza to pasta, and pasta to a salad. What is he to choose if he has to take exactly one of them? If Peter prefers the salad to the pizza, he should be happy to pay a bit of money to get the salad rather than the pizza. If he prefers the pizza to pasta, he should be happy to pay a bit to get the pizza instead of pasta. And if he prefers pasta to the salad, he should be happy to pay a bit to get pasta instead of the salad. It then seems that we can get arbitrary amounts of money from Peter by arguing him through the circle many times (Binmore 2009, pp. 13–14).

Since Peter’s preferences seem odd in this case, we can deem Peter irrational and confine ourselves to rational agents. The hope is that at least the preferences of *rational* agents can be represented in the way envisaged. Let us therefore assume an axiom that constrains the preferences of rational agents. A very efficient way to do so is to adopt the following axiom, which I call vNM1 due to the relation to the von Neumann–Morgenstern utilities (see Kreps 1988, Chap. 2).

*vNM1.* The preferences of a rational agent are asymmetric, i.e., no outcome is preferred to itself (we never have  $a \succ a$ ). The preferences of a rational agent are also negatively transitive, i.e., if  $a$  is preferred to  $b$ , then any outcome  $c$  will either be preferred to  $b$  or else will  $a$  be preferred to  $c$  (i.e., if  $a \succ b$  we have for all  $c$ :  $c \succ b$  or  $a \succ c$ ).

The strategy to confine oneself to *rational* agents and to lay down *axioms* for the preferences of rational agents is standard in decision theory (for simplicity I will often drop the qualification “rational” in “rational agent”). The strategy is to some extent innocuous because BDT is about the choices of rational agents only. The strategy has nevertheless a price because vNM1 is a substantive claim about rational agents. It purports to constrain the preferences of rational agents. That it does so may be denied, and much criticism of standard decision theory is based upon the charge that the theory misinterprets rational choice.

As a consequence, decision theory is really rationality-in rationality-out. It does not provide constraints of rationality for free. It is rather based upon substantive assumptions about rational choice. These assumptions are in a way translated into the maximization of expected utility.

For the following, it is useful to introduce more notations. We will say that outcome  $a$  is no less desired than  $b$ ,  $a \succeq b$  iff it is false that  $b \succ a$ . The agent will be said to be *indifferent* between outcomes  $a$  and  $b$ ,  $a \equiv b$  iff  $a$  is no less desired than  $b$  and  $b$  is no less desired than  $a$ . Our requirement of representation implies that

$$u(a) \geq u(b) \text{ iff } a \succeq b$$

and that

$$u(a) = u(b) \text{ iff } a \equiv b.$$

Assuming vNM1, we can obtain a first, very simple representation theorem, call it ORT (representation theorem with ordinal utilities; Kreps 1988, Chap. 3):

*ORT.* Assume that vNM1 holds and that the number of possible outcomes is finite or countable.

Then there is a mapping  $u$  from the outcomes to the real numbers such that (► 15.1) is satisfied.

ORT grants the existence of a representation. As the preferences of the rational agent impose a certain order on the outcomes, and as this order is preserved using the mapping  $u$ , the latter is called an *ordinal utility function*. Note though that the mapping and thus the representation is not unique. If  $u$  furnishes a representation, so does  $fu$ , the composition of  $u$  with any strictly monotonically increasing function  $f$  from the real numbers to the real numbers (Kreps 1988).

ORT underwrites the maximization of utility. For it is a platitude that rational agents choose what they prefer most (at least if we restrict the notion of rationality suitably). Given our representation, this is equivalent to saying that a rational agent chooses an option with a maximal utility.

All this is still far from BDT. We want to underwrite the maximization of *expected utility* in decisions such as Peter’s. To that end, von Neumann and Morgenstern and their followers introduce a new type of option. Let us start from a set of outcomes  $S$ . The new options are *lotteries*. If the agent chooses a lottery, a gambling device is used, for instance, a coin is flipped, and, depending on the result (heads or tails), a *prize* from  $S$  is realized. Each lottery only gives a finite number of prizes with a nonzero probability. Each outcome  $a$  from  $S$  can itself be regarded as a lottery that gives  $a$  for sure. Clearly, lotteries model options under which the outcomes are not fixed, but chancy.

The set of all lotteries with prizes in  $S$  has a certain structure; it is a mixture space. The essential point is that one can combine two lotteries to form a new lottery in the following way.

Assume that  $\mu$  and  $\lambda$  are lotteries and take  $\alpha$  to be a real number from the interval  $[0,1]$ . Under the new lottery, each outcome  $a$  from  $S$  has a probability of  $\alpha \times p_\lambda(a) + (1 - \alpha) \times p_\mu(a)$  if it has a probability of  $p_\lambda(a)$  under lottery  $\lambda$  and a probability of  $p_\mu(a)$  under lottery  $\mu$ . If the agent does not care how the values of the probabilities for the prizes arise, we can think of the new lottery as a two-stage lottery. In the first stage, a coin is flipped that lands heads with a probability of  $\alpha$ , and tails with a probability of  $(1 - \alpha)$ . If the coin lands heads, lottery  $\lambda$  is played in the second stage, whereas lottery  $\mu$  is played in the second stage if the coin lands tails. The new lottery is called a *compound lottery*; for obvious reasons, it is denoted as  $\alpha \times \lambda + (1 - \alpha) \times \mu$ .

Our question now is: What does it mean to choose rationally between lotteries? How are the preferences over lotteries constrained for rational agents?

It is first natural to assume that the agent's preferences over the lotteries obey the axiom vNM1. But we need two further assumptions to obtain expected utilities. The first is the *independence axiom*.

vNM2. Suppose that  $\lambda$ ,  $\mu$ , and  $\nu$  are lotteries and assume that  $\lambda$  is preferred to  $\mu$  (i.e.,  $\lambda \succ \mu$ ).

Then the compound lottery  $\alpha \times \lambda + (1 - \alpha) \times \nu$  is preferred to the compound lottery  $\alpha \times \mu + (1 - \alpha) \times \nu$ .

The independence axiom is not implausible (but see McClenen 2001). To see this, compare the alternative options  $\alpha \times \lambda + (1 - \alpha) \times \nu$  and  $\alpha \times \mu + (1 - \alpha) \times \nu$ . Let us think of them as two-stage lotteries with the same random process at the first stage. For instance, a coin is flipped that lands heads with a probability of  $\alpha$  and tails with a probability of  $(1 - \alpha)$ . If the coin lands tails, each of our two alternative lotteries gives  $\nu$ , in which case there is no difference between the lotteries. If the coin lands heads, the first lottery gives  $\lambda$ , whereas the second gives  $\mu$ . But  $\lambda$  is preferred to  $\mu$ . So it seems plain that the first lottery should be preferred to the second. The idea is that the ranking between the lotteries only depends on those cases in which the lotteries differ (see Machina 1982 for generalizations of expected utility theory without the independence axiom).

The last axiom is the *continuity axiom*.

vNM3. Suppose that  $\lambda$ ,  $\mu$ , and  $\nu$  are lotteries and assume that  $\lambda$  is preferred to  $\mu$  ( $\lambda \succ \mu$ ) and that  $\mu$  is preferred to  $\nu$  ( $\mu \succ \nu$ ). Then there is a real number  $\alpha$  strictly between 0 and 1 such that the compound lottery  $\alpha \times \lambda + (1 - \alpha) \times \nu$  is preferred to  $\mu$ . Likewise and conversely, there is another real number  $\alpha$  strictly between 0 and 1 such that  $\mu$  is preferred to  $\alpha \times \lambda + (1 - \alpha) \times \nu$ .

This is again not implausible. Suppose that  $\mu$  is ranked between  $\lambda$  and  $\nu$ , as is assumed in vNM3. Consider now compound lotteries of the type  $\alpha \times \lambda + (1 - \alpha) \times \nu$ . All of these lotteries can lead to  $\lambda$  and  $\nu$ ; however, their probabilities for  $\lambda$  and  $\nu$  differ. If  $\alpha = 1$ , the compound lottery coincides with  $\lambda$  and is thus preferred to  $\mu$ . Likewise, if  $\alpha = 0$ , the compound lottery coincides with  $\nu$ , and thus  $\mu$  is preferred to it. Let us now start with  $\alpha = 1$  and decrease  $\alpha$ . Thus, in the compound lottery, the higher ranked  $\lambda$  becomes less likely in favor of the lower ranked  $\nu$ . It is thus plausible to assume that the compound lottery becomes increasingly lower ranked for the agent. It is also plausible to think that, at some point, the compound lottery passes  $\mu$ . What the continuity axiom implies in this intuitive picture is this: There is a value of  $\alpha$  smaller than 1 such that the compound lottery is still preferred to  $\mu$ . Likewise, there is a value of  $\alpha$  larger than 0 such that  $\mu$  is already preferred to the compound lottery.

We can now state the *von Neumann–Morgenstern representation theorem* (vNMRT). Its first part, vNMRT.E, grants the *existence* of a representation:

vNMRT.E. Assume that vNM1 – vNM3 hold for a rational agent. Then there is a mapping  $u$  from the outcomes to the real numbers such that a lottery is preferred to another one iff its expected utility is larger than that of the second lottery.

As before, we obtain the expected utility of a lottery by adding the products of the utilities of the outcomes,  $u(a)$ , with their respective probabilities under the lottery.

vNMRT.E effectively introduces a new type of representation. A lottery is preferred to another iff its *expected* utility is higher. In this way, lotteries are evaluated using not their own utilities, but expected utilities. These derive from the utilities of the prizes and the probabilities. In this way, vNMRT.E underwrites the maximization of expected utility. To the extent that a rational agent conforms to vNM1 – vNM3 and chooses what she prefers most, she maximizes the expected utility.

The second part of the theorem, call it vNMRT.U, is about *uniqueness*.

vNMRT.U. The representation is unique up to affine transformations.

A transformation is affine if it is positive and linear. vNMRT.U implies that, if the utilities  $u(a)$  represent the agent's preferences, so do the utilities  $\alpha \times u(a) + \beta$ , where  $\alpha > 0$  and  $\beta$  are arbitrarily fixed real numbers. But vNMRT.U also grants that there are no other representations than these. Put differently, utilities do not have a zero point nor natural units. However, ratios of utility differences are uniquely fixed. This is necessary if the maximization of expected utility is to make sense. In  Table 15.2, for instance, it matters, not just which utility is larger than which, but also how utility differences compare to each other. It is important for the final choice that the difference between the utilities in the second column ( $100 - 190 = -90$ ) is  $(-9)$  times the difference between the utilities in the first column ( $200 - 190 = 10$ ). If the ratio of the differences were  $(-3)$  instead of  $(-9)$ , maximizing expected utility would favor the first option rather than the second. We have thus to make sense of utility differences being equal or being multiples of each other. This is achieved by the representation theorem by von Neumann and Morgenstern. It yields *cardinal* utilities.

The theory can easily be applied in practice (cf. Luce and Raiffa, Chap. 2; see below for a proviso). Consider Peter's decision with his car once more. There are four outcomes (car not breaking down and sold to the lawyer; car breaking down and sold to lawyer, etc.). As a rational agent, Peter should be able to rank these outcomes. Without loss of generality, he can set the utility of the highest ranked outcome at 1, and the utility of lowest ranked outcome at 0. For each other outcome  $a$ , he has to find a lottery that has only the highest and lowest ranked outcomes as prizes and that is equally ranked with  $a$ . If Peter's preferences satisfy the axioms, it is granted that there is such a lottery for each  $a$ . In this way, Peter will obtain a unique utility for each outcome. To determine what *option* to choose rationally, Peter has to think of the options as lotteries. For instance, the first option may be regarded as a lottery that provides one outcome with a probability of .8 and the other with .2. Peter should then choose the option that maximizes expected utility.

There is, however, a problem about the probabilities. In von Neumann–Morgenstern utility theory, they are simply put in by hand. They arise as parts of the lotteries that are presented as options. The probabilities are supposed to be intelligible (otherwise the lotteries would not be so), and they are assumed to obey the probability calculus. The theory does not say much as to

what the probabilities are and where their values come from. This is not very satisfying from a theoretical point of view.

It is often assumed that the probabilities are objective chances that characterize gambling devices such as roulette wheels. That is, there are mind-independent facts as to what the values of the probabilities are. The objective probabilities may quantify the strength of a disposition or be defined in terms of hypothetical frequencies. Under such an objectivist reading of the probabilities, von Neumann–Morgenstern theory only applies to lotteries that involve gambling devices or other systems for which an objective probability is defined. However, many real-world situations such as Peter's in our example are not like this. Here the probabilities are not objective, but rather reflect the agent's ignorance as to what outcome a choice may produce. We can only apply the von Neumann–Morgenstern theory to such cases if we grant that the axioms of the theory continue to hold for choices in which probabilities reflect the agent's ignorance (cf. Pfanzagl 1967). But can we take this assumption for granted?

Anscombe and Aumann (1963) provide an extension of von Neumann–Morgenstern utility theory to include subjective probabilities. They introduce two kinds of lotteries, viz. roulette lotteries and horse-race lotteries. Whereas the outcomes of the former have objective probabilities, the outcomes of the latter have not. Anscombe and Aumann consider compound lotteries that involve both roulette and horse-race lotteries. They assume that the agent's preferences over roulette lotteries obey the axioms of von Neumann–Morgenstern utility theory (or our axioms from above). Likewise, the agent has preferences over horse-race lotteries that have roulette lotteries as prizes. These preferences obey the axioms of von Neumann–Morgenstern utility theory as well. So we obtain two kinds of rankings and correspondingly two kinds of utilities – one for horse-race lotteries and one for roulette lotteries. Anscombe and Aumann make a few weak assumptions about how the two sorts of preferences are connected. Their main result is that the agent prefers one horse-race lottery to another iff a certain linear combination of the utilities of the prizes is higher. The coefficients that arise in this linear combination take the role that probabilities take in the assessment of roulette lotteries and can thus be interpreted as subjective degrees of belief.

Under Anscombe and Aumann's theory, a rational agent maximizes an expected utility that arises from subjective probabilities. Moreover, the subjective probabilities are not put in by hand, but are instead obtained from the preferences of the agent. So the theoretical problem about probability in the von Neumann–Morgenstern theory is to some extent avoided. However, the subjective probabilities only arise from preferences concerning gambling devices that can be characterized using objective probabilities in turn. In this sense, objective probabilities are conceptually prior. This allows for a certain criticism because the notion of an objective probability is sometimes taken to be problematic. The question then is whether we can do without objective probabilities altogether. The answer is positive and was in fact given prior to von Neumann and Morgenstern by Ramsey.

## Ramsey's Approach

Frank P. Ramsey's approach is contained in his paper "Truth and Probability" (Ramsey 1931), which was written about 1926 and first published in 1931. It did not become influential until much later, but it is no exaggeration to say that it contains the core of Bayesian decision theory.

Ramsey did not elaborate the details, but Suppes (1956) provides a logical reconstruction of Ramsey's proposal (see also Davidson and Suppes 1956; cf. also Bradley 2001).

Ramsey's aim is different from that of von Neumann and Morgenstern and in some way complementary to theirs. The latter want to provide foundations for utility theory, and their main idea is that utility represents the degree to which something is wanted. Ramsey's primary interest is in probability theory. He identifies probabilities (or at least some of them) with degrees of belief. Thus, probabilities *represent* degrees of belief. Each proposition  $q$  about which the agent has views is mapped to its probability  $p(q)$ . We require that  $p(q) > p(r)$  iff the agent is more confident about  $q$  than she is about  $r$ .

But what does it mean to say that one proposition is more strongly believed than another one? And what does it mean that a proposition  $q$  is believed to a degree  $p(q)$ ?

Ramsey's crucial idea is that degrees of beliefs are determined in virtue of the *effects* they would have on action. Note that Ramsey explains degrees of belief not just in terms of *actual* effects that belief states have on *actual* actions, but also using hypothetical effects a belief state would have, provided a rational agent faced a suitable choice.

Suppose, we want to know the degree to which Peter believes that his car will not break down in the next year. To find this out, we can ask him what he would do if he were faced with a choice between the following two options: Either selling the car to the lawyer or selling it to Mary. As things are in the example, the options boil down to something like (obtain \$3,000 from the lawyer if the car does not break down; obtain \$1,000 from him if it does) and (obtain \$2,000 and additional benefits from Mary if the car does not break down; obtain \$2,000 and fewer additional benefits if the car does do so). Peter is thus confronted with two different *bets* on the truth of the proposition that his car does not break down. Assume now that the utilities from  Table 15.2 specify how strongly the various outcomes are desired. Qualitatively, if Peter is almost certain that his car will not break down in the next year, he would give it to the lawyer if he faced the decision. If he is much less confident that his car will not break down, on the contrary, he would choose to give it to Mary. Thus, by confronting Peter with the hypothetical choice, we can constrain the strength of his confidence. As we shall see presently, we can extend this story to obtain quantitative degrees of belief. The more general point is that we can obtain probabilities, when we have a suitable set of outcomes with their utilities and when we consider the hypothetical choices that an agent would make if she were faced with certain bets. De Finetti's definition of subjective probabilities is based upon this idea (de Finetti 1931, 1937). His approach is obviously the counterpart of von Neumann–Morgenstern expected utility theory.

De Finetti's approach suffices to define probabilities, but there is a problem with it if our interest is in decision theory. What Peter would choose in a hypothetical choice is not only a function of his beliefs about the car, but also of the extent to which he desires the outcomes that can arise. We can only obtain probabilities in this way if utilities are provided. But how do we obtain the utilities? So far, we have only von Neumann–Morgenstern utilities, which were derived taking for granted some probabilities, and we are moving in a circle. What we should aim at then is to obtain utilities as measures of strength of desire and probabilities as measures of strength of belief *at the same time*. And the only facts that determine those strengths concern the hypothetical choices an agent would make if faced with suitable sets of options. The problem is that beliefs of various degrees and desires of various degrees cooperate to determine what the agent would choose. To obtain degrees of belief, we have to disentangle them from the

agent's desires with their various strengths. The virtue of Ramsey's approach is that it obtains both utilities and probabilities from hypothetical choices or preferences.

Assume an agent has preferences over a set of *outcomes* such that vNM1 is fulfilled. There are also *propositions* about which the agent has views and to which subjective probabilities will be assigned. Finally, the agent is faced with hypothetical choices between *bet-like options*, under which the outcome is a function of the truth of a proposition. These options need not involve gambling devices – the idea is only that, *given what the agent thinks*, various outcomes are possible for some option.

Ramsey assumes that some propositions are *desire-neutral*. The truth of such a proposition  $q$  makes never a difference as to how much an outcome is wanted. Thus, for any outcome  $a$ , the agent is indifferent between the options “ $a$  is realized and  $q$  (is true)” and “ $a$  is realized and  $q$  is false.” A good candidate for a desire-neutral proposition is the proposition that the number of people at Times Square at noon tomorrow is even. Whether this is so or not will not affect the degrees to which outcomes are desired by most people.

Ramsey's account can now be developed in three steps. The *first step* is to identify propositions that will obtain a degree of belief of .5. Suppose that the agent prefers outcome  $a$  to  $b$ . It thus matters to her whether  $a$  or  $b$  will obtain. Consider now a desire-neutral proposition  $q$ , its negation  $\neg q$  and the following options: “ $a$  if  $q$ ;  $b$  if  $\neg q$ ” and “ $b$  if  $q$ ;  $a$  if  $\neg q$ .” These are bets on the truth of  $q$ , but the bets differ because they connect the outcomes  $a$  and  $b$  to the truth of  $q$  in different ways. Ramsey proposes to say that the agent takes proposition  $q$  and its negation  $\neg q$  to be *equally probable* iff she is indifferent between the bets. In this case, the degrees of belief concerning  $q$  and  $\neg q$  are equal and they obtain the value .5.

This is exactly what an expected utility formalism would imply. If we know that  $p(q) = p(\neg q)$ , the expected utility of the first option is  $p(q) \times u(a) + p(\neg q)u(b) = p(q) \times (u(a) + u(b))$ , and the expected utility of the second option is  $p(q) \times u(b) + p(\neg q) \times u(a) = p(q) \times (u(a) + u(b))$ , which is the same. Since the utilities are identical, the agent is indifferent between the options.

The *second step* is to define when utility differences of two outcomes are equal. Consider the four outcomes  $a$ ,  $b$ ,  $c$ , and  $d$ . We want to say that

$$u(a) - u(b) = u(c) - u(d).$$

Not just is  $a$  preferred to  $b$ , and  $c$  to  $d$ , but also is  $a$  preferred to  $b$  to the same extent that  $c$  is preferred to  $d$ . But what exactly is this supposed to mean? Ramsey proposes the following condition: Let  $q$  be an desire-neutral proposition with  $p(q) = .5$ . The utility differences are identical and the equation holds iff the agent is indifferent between the following choices: “ $a$  if  $q$ ;  $d$  if  $\neg q$ ” and “ $b$  if  $q$ ;  $c$  if  $\neg q$ .”

This condition is again motivated by the idea that the preferences of a rational agent can be represented using expected utilities. If this is so, the agent is indifferent between two options iff the expected utilities are identical. The expected utilities of the options are  $p(q) \times u(a) + p(\neg q) \times u(d)$ , and  $p(q) \times u(b) + p(\neg q) \times u(c)$ , respectively. If they are identical, we have  $p(q) \times u(a) + p(\neg q) \times u(d) = p(q) \times u(b) + p(\neg q) \times u(c)$ . Since proposition  $q$  is desire-neutral, we have  $p(q) = p(\neg q)$ , and it follows that  $u(a) - u(b) = u(c) - u(d)$ , as stated above. Ramsey's condition makes only sense though if it holds for *any* desire-neutral proposition with  $p(q) = p(\neg q)$ . Ramsey introduces this as an assumption.

The *third* step, finally, is to fix the probabilities of arbitrary propositions. According to Ramsey, an agent believes a proposition  $q$  to a degree of  $p(q)$  iff the agent is indifferent between the following two options “ $a$  in any case” and “ $b$  if  $q$ ;  $c$  if  $\neg q$ ,” where the utilities for the outcomes  $a$ ,  $b$ , and  $c$  obey:  $(u(a) - u(c))/(u(b) - u(c)) = p(q)$ . To say this is again no more than reading expected utilities in the choices of the agent. Our agent is indifferent between the options mentioned iff the expected utilities are identical. This is equivalent to  $u(a) = p(q) \times u(b) + p(\neg q) \times u(c) = p(q) \times u(b) + (1 - p(q)) \times u(c)$ , where we assume that  $p(\neg q) = 1 - p(q)$ , as is familiar from the probability calculus. Thus, indifference is equivalent to  $p(q) = (u(a) - u(c))/(u(b) - u(c))$ , which is exactly the condition that Ramsey proposes.

In this way, we obtain probabilities that measure degrees of belief and utilities that measure the degrees to which outcomes are wanted by an agent. One can show that the degrees of belief obey the rules of the probability calculus. One can also obtain a representation theorem, according to which an agent prefers a bet-like option to another iff the expected utility is higher than the other. The representation of degrees of belief turns out to be unique, utilities are determined up to an affine transformation as before.

However, to obtain these results, some assumptions have to be made (see Ramsey 1931, pp. 74–75 and Suppes 1956). For instance, if the account is to work, an agent that prefers outcome  $a$  to  $b$  must prefer ( $a$  if  $q$ , and  $b$  if  $\neg q$ ) to  $a$ . An axiom that requires this may again be introduced as a constraint of rationality. Ramsey also needs a lot of outcomes that are desired to the same extent as are bet-like options. Axioms that guarantee this cannot plausibly be read as constraints of rationality, but rather concern the availability of suitable outcomes. Suppes (1956, p. 66) thus distinguishes *axioms of rationality* from *axioms of structure*. The latter constrain possible outcomes and options. As a consequence, the theory is not only restricted to rational agents but also to decisions for which the axioms of structure hold.

How far does this get us concerning BDT? At first glance, Ramsey reconstructs an agent’s preferences or choices using an expected utility formalism to obtain numerical degrees of beliefs and utilities. From this perspective, expected utility maximization seems an input to his theory (cf. Kaplan 1996, p. 168). Some remarks of Ramsey’s suggest such a reading too. As he puts it, his approach “is based throughout on the idea of mathematical expectation” (p. 70). Even under this interpretation, we have a remarkable achievement because it is not at all obvious that BDT can be used to define degrees of beliefs and to measure utility differences. But Ramsey’s result would not justify BDT. In this spirit, Kaplan (1996, p. 176) complains that Ramsey’s results do not underwrite the idea that rational agents maximize expected utility.

However, there is indeed much more to Ramsey’s results. Forget about the *motivations* for the three steps and take them to be *definitions* of degree of belief and degree to which an outcome is wanted, or of probability and utility. The definitions immediately imply that, in certain simple choices, the agent maximizes expected utility (this is just how the definitions are devised). What Ramsey shows is that probability and utility can always be defined in this way given that some axioms of rationality and of structure hold. It can also be shown from the axioms that expected utility is maximized more generally and not just in those choices that are considered in the definitions. Thus, granted the axioms, it follows that rational agents maximize some expectation value. If Ramsey’s definitions capture our pre-theoretic knowledge about belief, desire, and their degrees, the expectation value reflects the beliefs and desires and their strengths, and BDT is justified.

That BDT follows from assumptions that do not explicitly require the maximization of expected utility is even clearer from Savage’s theory.

## Savage's Theory

Leonard J. Savage's "Foundations of Statistics" (1954/1972) is often regarded as *the* articulation and defense of Bayesian decision theory. Fishburn (1979, p. 191) calls it "[t]he most brilliant axiomatic theory of utility ever developed" (he does not mention the Bolker-Jeffrey theory). Savage's theory draws on work by Ramsey, de Finetti, and von Neumann and Morgenstern (for earlier work by Savage see Friedman and Savage 1948, 1952).

Savage's conceptualization of choices is markedly different from what we have seen so far. The role of the conditions (whether the car will break down or not) is taken by states and events. Each *state* corresponds to a full description of all the objects the agent is concerned with (i.e., of the *world* in Savage's terms). *Events* are sets of states. An event in our example is that the car breaks down in the next year. This is not a state, because it leaves many things open, for instance, whether Peter will be invited to Mary's parties. In Savage's terms, all states in which the car breaks down constitute the event that the car breaks down.

A state in Savage's sense is not the state of the world at the moment of choice, but also encompasses future developments, for instance, whether the car will break down in the next year. The agent does not know what the real state is and whether certain events obtain. Savage's theory will eventually yield a probability function for the events (the probability function is defined on *events* rather than on *states* because there may be so many states that only suitable sets of states have probabilities).

An outcome is called a *consequence* by Savage. This term is a bit misleading because Savage's consequence means a set of consequences in common parlance. Each option is conceptualized as an *act*, which is a function from states to consequences. For instance, if Peter sells his car to Mary, the state "the car does not break down in the next year and Mary hosts many parties next year" is mapped to the consequence "Peter obtains \$2,000, Mary drives the car and Peter is invited to all of Mary's parties"; the state "the car breaks down in the next year and Mary hosts many parties next year" is mapped to the consequence "Peter obtains \$2,000, Mary drives the car for some while, Mary becomes angry with Peter and Peter is only invited to one of Mary's parties"; and so on.

Since acts are functions, an act together with a state fixes the consequence. The probabilities for the events can thus not depend on the act chosen. This may seem problematic. As was suggested above, the probability that the car breaks down may depend on whether it is sold to the lawyer or to Mary. Such a dependence is not possible in Savage's theory. If the car's breaking down is an event, there can only be one probability for it. Nevertheless, Savage's theory can deal with the example if the decision problem is conceptualized using different states. The idea is that each state does not contain information about whether the car will break down, but only information about whether the car will break down *if* Mary owns it, and so on. The event that the car will break down *if* Mary owns it, has as fixed probability, as Savage requires, but choice of the agent will effectively make a difference for the probability that the car breaks down.

The development of Savage's theory is quite different from that of Ramsey's. First, Savage does not need desire-neutral propositions. Second, he obtains utilities *after* probabilities, whereas Ramsey goes the other way round.

Savage starts from *preferences over acts* (where acts correspond to options in our earlier terms). Savage's first assumption (S1) is that vNM1 is satisfied for preferences over acts.

Second, Savage assumes the famous *sure thing axiom*, which is not far from the independence axiom from above (cf. Fishburn and Wakker 1995). Here is a very rough formulation:

S2. If we compare two acts concerning the extent to which they are desired, states on which the acts agree do not matter.

In more detail, suppose that an act  $f$  is preferred to another act  $g$ , but that  $f$  and  $g$  (as functions on the states) coincide for some event. That is, if the event obtains, the acts yield exactly the same consequence. For instance, the acts  $f$  formalizing “buy lottery ticket no. 1 for \$1” and  $g$  formalizing “buy lottery ticket no. 2 for \$1” do not make any difference if both tickets lose. Suppose now that both acts are modified because the regulations for lotteries change, and each customer gets 1 cent back if her ticket loses. We then have new acts  $f'$  and  $g'$ , which still agree on the event that both tickets lose. The *sure thing axiom* requires that  $f'$  is preferred to  $g'$ .

Taking for granted this axiom, Savage can define the notion of “one act being preferred to a second one, *given event B*.” Intuitively, this captures the idea that a certain act is preferred to another *for certain circumstances*. For instance, given that the car breaks down, Peter may prefer selling his car to Mary to selling it to the lawyer.

Savage’s third axiom concerns what he calls *constant acts*. An act is constant iff it has the same consequence for every state. The preferences over acts include preferences over constant acts, and these induce preferences over consequences in a straightforward way. Savage’s third axiom reads as follows.

S3. If a constant act is preferred to another constant act, then it is also preferred to the latter given an arbitrary event.

Intuitively, if one act always has the same consequence  $c_1$ , if another act always has the same consequence  $c_2$ , and if we prefer the first act to the second, this does not change if we acquire new information on the real state of the world. One may object that there are counterexamples. For instance, I may prefer taking the bus to work (constant consequence: pay \$2, 10 min time for getting to work) to walking (constant consequence: pay \$0, 20 min for getting to work) quite generally, but not if I learn that there is rain. However, the counterexample relies on the fact that certain aspects of a walk that matter to me are not taken into account in describing the consequences. If we properly take into account what matters to me, walking to work is not a constant act and the counterexample disappears.

The next step is to think about acts that take the role of bets. I will call an act a bet on an event (with prizes  $c_1$  and  $c_2$ ) if the act maps each state from the event to consequence  $c_1$ , and each state outside the event to another consequence  $c_2$ , where  $c_1$  is preferred to  $c_2$ . Real bets can be conceptualized as bets on events, but other options may be bets too. To obtain any bet on an event, Savage has to assume that

S4. There is some consequence (or some constant act) that is preferred to another consequence.

Savage furthermore assumes the following axiom.

S5. If some bet on event  $A$  with consequences  $c_1$  and  $c_2$  is preferred to some bet on event  $B$  with consequences  $c_1$  and  $c_2$ , then every bet on  $A$  with consequences  $c_3$  and  $c_4$  is preferred to every bet on  $B$  with consequences  $c_3$  and  $c_4$ .

This axiom requires consistency in accepting bets on events. Suppose, I have to choose between the following bets. The first bet pays \$1 if  $A$  occurs, and \$2 if not. The second bet pays

\$1 if  $B$  occurs, and \$2 if not. Suppose, I prefer the first bet to the second. Let us now modify both bets by slightly changing the consequences. A prize of \$1 is consistently replaced by one of \$100, and a prize of \$2 is consistently replaced by one of \$500. According to S5, I should continue to prefer the first to the second bet.

This seems plausible. A rationale may be that my reactions to the bets from the example should only depend on whether  $A$  seems more likely to me than  $B$ . If I take  $A$  to be more likely than  $B$ , it will always be preferable to bet on  $A$  if the prizes do not differ. However, Savage does not yet have the notion of probability. He rather argues the other way round. He uses the axioms S1–S5 to make sense of *qualitative* degrees of belief. A rational agent takes event  $A$  to be more probable than event  $B$  iff she prefers a bet on  $A$  with consequences  $c_1$  and  $c_2$  to a bet on  $B$  with consequences  $c_1$  and  $c_2$ . Axiom S5 implies that this will be so for every pair of consequences  $c_1$  and  $c_2$ . It turns out that the relation “... is taken to be more probable than ...” has many features that one would expect for probabilities.

However, Savage's aim is to derive *quantitative* probabilities. We need a function  $P$  on the events such that  $p(A) > p(B)$  iff  $A$  is taken to be more probable than  $B$  in the sense just defined.

To obtain such a representation that is furthermore unique, Savage introduces another axiom. It is more technical and reads as follows:

S6. If one act is preferred to a second one and if an arbitrary consequence  $c$  is given, the set of states can be partitioned into tiny sets that each have the following property: If the preferred act is slightly modified as to yield  $c$  on the tiny set and otherwise left unaffected, it is still preferred to the second act. An analogous statement holds about the second act.

In Suppes's terms, this is at least in part an axiom of structure. Savage suggests the following strategy to justify S6: Suppose, you prefer climbing to swimming at a particular occasion. Assume furthermore that you do not like a certain pain. Consider now a long series of  $n$  flips of a coin you think to be fair, where  $n$  is a natural number. Modify climbing such that you suffer the pain if one particular series of heads and tails occurs and the consequence of climbing otherwise. Of course, the higher  $n$  is, the less likely will a particular series of heads and tails be regarded. What S6 claims is that there is some number  $n$  such that you would still prefer modified climbing to swimming.

From S1–S6, Savage obtains a unique representation of probabilities. Given a notion of probability, it is not difficult to obtain utilities, as we know from the von Neumann–Morgenstern theory. The crucial step is to introduce acts that are counterparts to lotteries in the von Neumann–Morgenstern framework. A lottery has finitely many prizes (consequences in Savage's terms), but they are now connected to events on which subjective probabilities are defined.

Savage can show that the preferences over such lotteries follow the expectation value of the consequences. To extend this result to arbitrary acts, he needs an additional axiom.

S7. Suppose, every consequence that an act  $f$  yields for event  $A$  is preferred to another act  $g$  given  $A$ , then  $f$  is preferred to  $g$  given  $A$ . (And an analogous statement holds, in which “preferred” is replaced by “is not preferred to”).

Assuming S1–S7, we do not only obtain a unique quantitative probability that represents qualitative probability, but also the existence of a utility function such that one act is preferred to another one iff the expected utility of the consequences is larger for the first act than for the second. That is, rational agents to which S1–S7 apply follow MEU. Their preferences can be represented using expected utilities.

## The Bolker–Jeffrey Theory

---

Although Savage's theory is the most famous version of Bayesian decision theory, there is an interesting alternative, viz. the theory developed by Ethan D. Bolker and Richard C. Jeffrey (Jeffrey 1965a, b; Bolker 1966, 1967). While the main mathematical results upon which the theory is based are due to Bolker, Jeffrey is the most prominent philosophical defender of the theory.

In each of the approaches reviewed so far, probabilities and utilities attach to different kinds of entities. Probabilities are assigned to propositions (Ramsey) or to events (Savage), whereas utilities characterize outcomes or acts. In the Bolker–Jeffrey theory, on the contrary, both probabilities and utilities attach to propositions. The underlying idea is that both belief and desire are *propositional attitudes*, i.e., attitudes with a content that can be expressed using a proposition. That belief is a propositional attitude is uncontroversial, but desires may seem to aim at objects rather than at propositions. For instance, we say that Peter wants a new bike. However, what we really mean by this is that Peter wants to own a new bike or to drive a new bike and this can be expressed using a proposition. Another problem is that we can only desire something that does not yet obtain. Jeffrey does not have a rebuttal to this and suggests that his theory is really about desirability from the agent's point of view. The idea is that some propositions would count as better news to the agent than others (Jeffrey 1965a, p. 72). Actions and options are also conceptualized as propositions, viz. as those propositions that describe the actions or the options.

In Jeffrey's framework, the agent has two attitudes to each proposition: She takes the proposition to be probable to some extent, and she takes the proposition to be desirable to some extent. Let us say that proposition  $q$  is preferred to proposition  $r$  iff the agent takes  $q$  to be more desirable than  $r$ . As before, we want a representation, i.e., a function that maps any proposition  $q$  to a real number  $u(q)$  such that  $q$  is preferred to  $r$  iff  $u(q) > u(r)$ . In the same way, we want a function that maps any proposition  $q$  to a real number  $p(q)$  such that  $q$  is taken to be more likely than  $r$  iff  $p(q) > p(r)$ . Jeffrey assumes that the preferences over the propositions are asymmetric and negatively transitive.

A proposition does not uniquely fix the state of the world. The proposition  $q$  that Peter sells the car to Mary is compatible both with the car breaking down and with the car not breaking down in the next year. Obviously, the desirability of  $q$  should be responsive to the desirabilities of both possibilities. If both the propositions  $q_1$  (that Peter sells the car to Mary and that the car breaks down) and the proposition  $q_2$  (that Peter sells the car to Mary and the car does not break down) are not very desirable,  $q$  cannot be either. But the desirability of  $q$  should also depend on how likely  $q_1$  and  $q_2$  are taken to be. If  $q_1$  is very desirable but regarded very unlikely, then it should not have a large impact on the desirability of  $q$ . This suggests the following formula:

$$u(q) = \frac{u(q_1) \times p(q_1) + u(q_2) \times p(q_2)}{p(q_1) + p(q_2)}$$

Here we assume that  $q_1$  and  $q_2$  are incompatible, but jointly exhaust  $q$ . According to the formula,  $u(q)$  is a *conditional expectation value*. It equals the expectation value of the desirability of the various possibilities compatible with  $q$ , given that  $q$  is true. The less likely a possibility is given  $q$ , the smaller is its contribution to the desirability of  $q$ .

This may seem plausible, but the question is again how we can justify the formula. Here is where Bolker's mathematical results step in. We have the following representation theorem:

Suppose, an agent has preferences on a set of propositions with a certain structure. Assume furthermore that the preferences are asymmetric and negatively transitive and that they obey further constraints that I will not list here. Then there is a utility function  $u()$  on the propositions that represents desirability as required. Furthermore, there is a probability function on the propositions such that, for every proposition  $q$ ,  $u(q)$  can be expressed as the conditional expectation value of the utility, given  $q$ . The representation is unique up to some transformations. Interestingly, these are not just affine transformations of the utilities, and the probabilities are not uniquely fixed (see Bradley 1998 and Joyce 1999, Chap. 4 for attempts to restore the uniqueness of the probabilities in the Bolker–Jeffrey framework).

The Bolker–Jeffrey theory has some advantages over Savage's (see e.g., Eells 1982, pp. 82–86). For instance, the probabilities for various outcomes can now depend on the action chosen. This is so because the *conditional* expectation value is taken. It is also interesting that the difference between utility and expected utility disappears in the theory – each utility can be written as a conditional expectation value.

## Current Research: Bayesian Decision Theory and Risk

Standard decision theory of any brand seems to give unequivocal advice for risky choices. The advice is of course to maximize expected utility. This advice is justified by representation theorems according to which rational agents maximize expected utility. To be sure, since decision theory only captures a narrow notion of rationality, its theorems themselves do not ground a *moral* evaluation. But it is plausible to assume that moral preferences and moral choices should respect the constraints from decision theory too (see Dreier 2004 and Lumer 2010 for the relation between decision theory and morality).

However, intuitively, it does not always seem rational to maximize expected utility. Consider a choice between the options represented in  [Table 15.3](#).

The numbers in the upper halves of the boxes specify the utilities, whereas the numbers in the lower halves denote the probabilities as before. Let us grant that there cannot be any doubts about the probabilities – they may be objective probabilities over outcomes of gambling devices, or subjective probabilities based upon the available evidence. We observe that option  $O_2$  has a slightly higher expected utility (160) than  $O_1$  has (150) and that  $O_2$  is thus recommended as rational by standard decision theory. But many people will prefer  $O_1$  to  $O_2$  because  $O_2$  seems too risky.  $O_2$  has a higher expected utility than  $O_1$ , but only because an

 **Table 15.3**  
Another decision problem

	$q$	$\neg q$
$O_1$	200	100
	.8	.2
$O_2$	0	16,000
	.99	.01

The numbers in the upper halves of the boxes denote utilities

outcome with an extremely high utility has a tiny probability. If  $O_1$  is chosen, by contrast, there will at least be a utility of 100 for sure. Is it really irrational to prefer  $O_1$  to  $O_2$ ?

One can, of course, simply deny this by rejecting standard and particularly Bayesian decision theory. I will turn to related criticism of the theory in the next section. In this section, my question is only what Bayesians themselves can say.

Before discussing our question, it is useful to mention that, in decision theory, “risk” has often a technical meaning that does not reflect ordinary usage (cf. Hansson 2011). First, decisions are said to be under risk iff probabilities are available to the agent. This is contrasted to decisions in which the agent has no probabilities for the outcomes. The terminology dates back to Knight (1921).

Second, in the literature, expected utility is sometimes multiplied with a minus sign and called risk. In these terms, Bayesian decision theory implies that rational agents minimize risk (e.g., Berger 1980, p. 8). The pertinent meaning of “risk” is at odds with much common usage of “risk,” where risk is either quantified as the probability of some bad event provided a specific option is chosen, or as the expected harm that a choice produces (where potential benefits are not taken into account, see Hansson 2011).

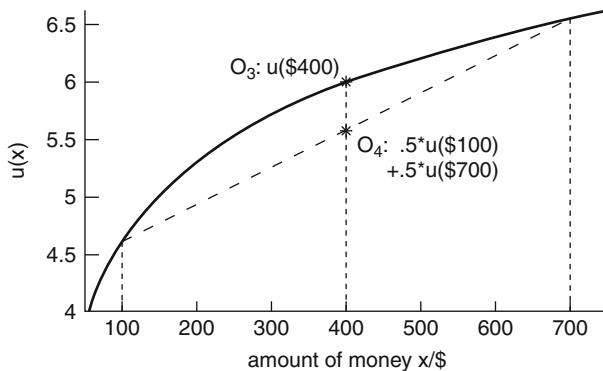
Finally, there is a technical notion of risk aversity to which we will turn presently (see Friedman and Savage 1952; Pratt 1964; Arrow 1970).

In our example, Bayesians will certainly insist that it is not rational to prefer  $O_1$  to  $O_2$  if the utilities and probabilities are as in [Table 15.3](#). To prefer  $O_1$  to  $O_2$  is to have preferences that do not obey the axioms of the theory. The axioms are partly axioms of structure, partly axioms of rationality. If axioms of structure are violated, nothing is wrong with preferring  $O_1$  to  $O_2$ , but it is unlikely that axioms of structure are violated (see below). So presumably, axioms of rationality are flouted and  $O_1$  cannot be the rational option to choose.

If an agent insists on preferring  $O_1$  to  $O_2$ , one can doubt that the utilities in the boxes properly reflect the agent’s utilities to begin with. In decision theory, the utilities are constructed from the agent’s preferences. If  $O_1$  is preferred to  $O_2$  and if the axioms are obeyed, we will never obtain utilities as in the boxes of the table.

Proponents of standard decision theory have additional resources to explain why we might want to prefer  $O_1$  to  $O_2$  in the example, while not rejecting the axioms of decision theory. The fault may simply be that we do not properly interpret the utilities from the table. We tend to “translate” the utilities in something homely by using some good as a proxy. For instance, in the example, we identify a utility of 100 with \$100, a utility of \$10,000 with \$10,000 and so on. But this is a mistake because the utility of money is in fact not a linear function of the amount of money. This is the main upshot of the St. Petersburg paradox. Since the utility of additional \$100 decreases, as the amount of money increases, we underestimate a utility of 10,000. We compare it to the utility of \$10,000, which is maybe not 100 times the utility of \$100, but only two times that utility.

In other words, if the numbers in the table do not reflect utilities, but rather some other good (e.g., money), an expected utility formalism may well imply that we should rationally choose the option that looks less risky. To see this, assume that the utility  $u(x)$  of some money  $x$  is a concave function of the amount of money  $x$ . That is, the utility increments for an additional \$1 decrease, as the amount of money grows (we still assume that the utility is a monotonically increasing function of the amount of money). The logarithm is an example of a concave function (cf. [Fig. 15.1](#)). Compare now the following two options ([Table 15.4](#)): The first option ( $O_3$ ) gives the agent \$400 for sure, the other ( $O_4$ ) is a gamble between \$100 and \$700.

**Fig. 15.1**

**Utility as a function of the amount of money.** The function is concave and thus illustrates risk aversity

**Table 15.4**

#### A yet other decision problem

	p	$\neg p$
$O_3$	\$400	\$400
	.8	.2
$O_4$	\$100	\$700
	.5	.5

Intuitively,  $O_4$  is more risky, but its *expected monetary value* is the same as that of  $O_3$ . Consider now the *expected utilities* of the options. We obtain for  $O_3$

$$u(\$400)$$

and for  $O_4$

$$.5 \times u(\$100) + .5 \times (\$700).$$

As  $u(x)$  is a concave function (see Fig. 15.1), the expected utility for  $O_4$  is less than that for  $O_3$ . Consequently, the agent should prefer  $O_3$ , which is the less risky option.

In this way, standard decision theory can rationalize preferences for less risky options. Utility functions that are concave (in some domain) as is  $u()$  in our example are said to capture *risk aversity* (in that domain). Many utility functions for goods are presumably concave in a large domain and thus encapsulate risk aversity. However, standard decision theory can only underwrite the choice of less risky options if risk is assessed in terms of other goods (e.g., money). An agent can never be risk averse concerning utilities (see Pratt 1964 and Arrow 1970, particularly Chap. 3, for risk aversity).

It may be objected that some people do not like (or, alternatively, enjoy) the uncertainty associated with real lotteries. This suggests that some prefer a *utility* of 400 for sure to a gamble with prizes of utilities 100 and 700, respectively (or the other way round). However, if the uncertainty of

an outcome is perceived to be bad or good in itself by the agent, the decision problem has to be redescribed. In this way, a contradiction with standard decision theory can be avoided.

Nevertheless, some ways of handling risk that do not seem irrational cannot be described in an expected utility framework. This is so when somebody applies the Maximin Rule, while having probabilities for each outcome given an option. See Kahneman/Tversky (1979), Kavka (1980), Ayres/Sandilya (1986), and Ekenberg et al. (2001) for further discussion.

## Further Research

---

To conclude, I will flag the most controversial issues raised by standard decision theory. I focus on more philosophical issues and do not review purely formal work in decision theory (see Fishburn 1979, 1982; Hammond 1998a, b; Sugden 2004).

1. *Is the maximization of expected utility a necessary condition of rational choice?* The main claim of standard decision theory is that rational agents maximize expected utility. As the maximization of expected utility can be derived from axioms, the question boils down to whether the axioms put down necessary conditions of rational choice and whether the axioms of structure are applicable. The standard view about the axioms of structure is that they may be fulfilled in most cases of interest at least by suitably redescribing the choice under consideration. The main issue then is whether the axioms of rationality reflect genuine constraints of rationality. This is the subject of much controversy (see e.g., Sugden 1985). Many objections against standard decision theory are cast in terms of paradoxes (e.g., Allais 1953; Ellsberg 1961). The paradoxes present choices between options. The choices that are intuitively judged to be rational by most people violate the requirement of maximizing expected utility. The paradoxes are discussed in [Chap. 19, Paradoxes of Rational Choice Theory](#), in this volume. Other criticism against standard decision theory grows out of empirical research (see below).

There are also attempts to justify some constraints of rationality from standard decision theory. So-called Dutch Book-arguments are most famous in this respect. They show that an agent who violates certain requirements is committed to accept sets of bets under which she will lose no matter what events materialize. There are well-known “Dutch-Book” arguments for the axioms of the probability calculus (Ramsey 1931, see also Gillies 2000, Chap. 4). It is controversial though whether agents that would accept Dutch books are irrational (see Kadane 1996 for a critical perspective).

2. *Is the maximization of expected utility a sufficient condition of rational choice?* As I have stressed above, standard decision theory focuses on a single choice and concentrates on a core notion of rationality to start with. Thus, other constraints of rationality seem possible. However, since the formalism of standard decision theory is thought to be extremely powerful, some may feel tempted to say that every requirement of rationality should be of the kind we have seen in developing the theory. If this is right, then maximizing expected utility may be sufficient for rational choice.

It is nevertheless very plausible that there are additional constraints of *epistemic/theoretical* rationality on the *probabilities* or the beliefs of the agent. As we have seen, the axioms of Bayesian decision theory (Ramsey, Savage, and Jeffrey) imply that the probabilities of an agent fulfill the axioms of the probability calculus. These axioms do not fix the

values of most probabilities though and only constrain combinations of them. But probabilities should also reflect the available evidence. If a coin has landed heads in about 50% of its flips so far, we should assign it a probability close to .5. Proponents of Bayesian decision theory reconstruct this by requiring that rational agents update their beliefs using empirical evidence. When data  $d$  are observed, the agent's probability for proposition  $q$ ,  $p(q)$ , should be replaced by the *posterior* probability  $p'(q) = p(q|d)$ , i.e., the probability for  $q$ , given the data. This is called *Bayesian conditionalization*. The posterior probability  $p(q|d)$  equals  $p(d|q)p(q)/p(d)$  via Bayes' Theorem. Bayesian conditionalization is sometimes justified by a "Dutch Book"-argument (Teller 1973; Lewis 1997), which is, however, very controversial (e.g., Howson 1995, pp. 8–10). Bayesian conditionalization is generalized to cases in which the data are uncertain via Jeffrey conditionalization (e.g., Jeffrey 1965a, Chap. 11).

As is manifest from Bayes' Theorem, Bayesian updating makes the posterior probability depend on the *prior* probability  $p(q)$ . This dependence is often "washed out" due to conditionalization. Under certain restrictions, the posteriors of agents with different priors, but with the same *likelihoods*  $p(q|d)$  converge, as the agents update their beliefs using the same data (Savage 1954/1972, pp. 46–50, Blackwell and Dubins 1962). It remains true though that finite data do not uniquely fix the values of the probabilities in this framework. This is often thought to be a problem since there seem to be cases in which some value of a probability seems eminently more reasonable than others.

*Objective Bayesians* claim that there are additional constraints on the values of probabilities (Williamson 2009, 2010). They typically demand that the priors obey the *Principle of Insufficient Reason* (or the Principle of Indifference; see Keynes 1921, p. 42 for a statement). Consider a set of mutually inconsistent, but jointly exhaustive propositions. According to the principle, each proposition should receive the same probability unless there are any reasons that speak in favor of one proposition rather than its rivals. This principle is generalized by the MAX-ENT principle by E. T. Jaynes (1957, 1968, 1979). However, the Principle is fraught with difficulties. The problem is that there are often many alternative ways to partition the space of possibilities, and various partitions will lead to different probability assignments. In such situations, the Principle cannot be applied unless one partition can be shown to be the right one (Gillies 2000, pp. 37–49).

To sum up, it is agreed among Bayesians that there is at least one other constraint of rationality (Bayesian updating). It affects degrees of belief only, but it is arguable that a choice is irrational in a broader sense if it draws on probabilities that have not been updated properly (see Earman 1992 for Bayesian updating and the implications for the philosophy of science).

3. *Can the maximization of expected utility guide choices?* As presented, decision theory is normative, since it constrains the choices of rational agents and since rationality is a normative notion. Now some normative principles such as moral principles can *guide choices*. That is, they give substantive answers to the practical question "What shall I do?" when it is asked from a first-person perspective. The question then is whether the maximization of utility can function in the same way. This is unlikely, because expected utility maximization is empty, unless utilities and probabilities are given, and because Bayesians only obtain utilities and probabilities from the agent's own preferences and choices. We cannot expect that these preferences are already given to an agent who deliberates what to do. Consequently, maximization of expected utility cannot guide choices

(see Peterson 2008, Chap. 2 for a recent version of this criticism, cf. also Bermúdez 2009, see also [Chap. 16, A Philosophical Assessment of Decision Theory](#)).

Bayesians have two kinds of replies. One reply is that guiding action is not an aim of the theory. As said above, the maximization of expected utility is a development of means-ends rationality, and means-ends-reasoning does not guide choices if the agent has not yet decided about her preferences. On this view, Bayesian decision theory expounds requirements of consistency on preferences. The second reply is that Bayesian decision theory can guide actions if combined with additional assumptions. For instance, if we assume a function for the utility of money, we can obtain more substantial advice. On this view, Bayesian decision theory would resemble the axioms of Newtonian mechanics. The latter have only observable consequences if they are combined with empirical hypotheses concerning forces.

4. *When should standard decision theory be applied? And are Bayesians committed to think that every decision is a decision under risk in Knightian terms* (Knight 1921)? Decision theory, as it has been presented, is concerned with one single decision. One can however apply the theory to a series of decisions. The basic idea is to regard a series of “small” decisions as one “big” decision. The options of this big decision are plans how to decide in the series of the little decisions. The plans may be conditional on information that only becomes available as the plans are carried out. Savage (1954/1972) famously restricts this strategy to what he called small worlds. Roughly, small worlds are such that we prefer the principle “Look before you leap” to the principle “You can cross that bridge when you come to it” (Savage 1954/1972, 16). Binmore (2009) makes a proposal on how to extend the scope of Bayesian decision theory to “bigger” worlds.

When we focus on a single choice, there is a very simple answer to the question when standard decision theory applies. It applies to those kinds of situations in which the axioms of structure hold. It is sometimes suggested that Bayesians take every decision problem to be one under risk in the terms of Knight and that they always recommend the maximization of expected utility. What is right about this is that Bayesians need not assume there to be objective probabilities to maximize expected utility. However, if the axioms of structure do not hold (or if the other axioms do not impose necessary conditions on rationality in the case at hand), then Bayesians are not committed to maximizing expected utility.

5. *Can one use Bayesian decision theory for descriptive tasks?* Although we have hitherto characterized decision theory as normative, the theory may also be used for descriptive purposes. If we assume that some agents are indeed rational in the sense of the axioms, their choices can be described using MEU. Whether agents are indeed rational is a matter of contingent fact. Empirical research is often taken to speak against the hypothesis that real-world agents maximize expected utility (see e.g., Kahneman/Tversky 1979 and Tversky/Kahneman 1981; consult Camerer 1995 for an overview; see also [Chap. 21, Real-Life Decisions and Decision Theory](#)). Related claims have to be taken with care though because standard decision theory is difficult to falsify. Typically, additional assumptions have to be made to obtain testable predictions from decision theory.

The normative claims from standard decision theory are, of course, not falsified if real people do not maximize expected utility. We cannot infer what we should do by looking at what people do in fact do. It would, however, be strange if most people were not rational for most of the time. Empirical results have thus some indirect impact on the fate of standard decision theory as a normative theory and they have in fact inspired alternatives (see e.g., Loomes and Sugden 1982 and Weirich 2004).

When decision theory is applied for descriptive purposes, rationality effectively drops out of the picture. It is immaterial for descriptive purposes whether a set of axioms are in fact constraints of rationality or not. The only question is whether the axioms are fulfilled for real-world agents. This is the basis for a more pragmatic approach to decision theory. The idea is that the axioms *define* a technical notion of rationality and that the theory can be applied to those agents that are rational in the technical sense (cf. Kreps 1988, pp. 4–6).

6. *What does Bayesian decision theory imply about human motivation?* Theories of motivation provide a framework for explaining actions and choices. According to the Humean theory of motivation (Smith 1987), actions are properly explained in terms of belief-desire pairs. Humean explanations in terms of beliefs and desires rationalize an action in a minimal sense (Davidson 1963). Anti-Humeans often claim that at least some actions are more appropriately explained using beliefs only (for instance, moral beliefs; cf. Nagel 1970). There is some affinity between Bayesian decision theory and the Humean theory of motivation because the former is often cast in terms of beliefs and desires and because many Bayesians are in fact committed to something like the Humean theory of motivation. Humeans, on the contrary, need a clear-cut distinction between beliefs and desires and may resort to standard decision theory to this end. However, Bayesian decision theory does not provide an independent argument for the Humean theory of motivation. It is also compatible with other views. It is arguable that the axioms do not properly apply to desire, but rather to other states (e.g., judgments of desirability or moral preferences; see Lewis 1988, 1996 and Price 1989 for discussion).
7. *What kind of probabilities ground rational action?* This chapter has concentrated on *evidential* decision theory in which the expected utilities arise from conditional probabilities in the most general case (think of the example in the introductory section and of the Bolker–Jeffrey theory for the moment). Proponents of *causal* decision theory, on the contrary, claim that the probabilities should not be conditional ones, but rather reflect causal knowledge. See Gibbard/Harper (1981), Lewis (1981), Eells (1982), Sobel (1994), and Joyce (1999).

Responding to some real or perceived problems of standard and particularly Bayesian decision theory, economists, philosophers, and others have developed variants (e.g., Kaplan 1996; Peterson 2008) or alternatives (e.g., Kahneman/Tversky 1979; Loomes and Sugden 1982). The discussion about Bayesian decision theory is alive as ever.

## References

---

- Allais M (1953) Le comportement de l'homme rationnel devant le risque: critique des postulats et axioms de l'école Américaine. *Econometrica* 21:503–546
- Anscombe FJ, Aumann RJ (1963) A definition of subjective probability. *Ann Math Stat* 34:199–205
- Arrow K (1970) Essays in the theory of risk-bearing. North-Holland, Amsterdam
- Ayres RU, Sandilya MS (1986) Catastrophe avoidance and risk aversion: implications for formal utility maximization. *Theory Decision* 20:63–78
- Barberà S, Hammond PJ, Seidl C (1998) Handbook of utility theory, vol I, Principles. Kluwer, Dordrecht
- Barberà S, Hammond PJ, Seidl C (2004) Handbook of utility theory, vol II, Extensions. Kluwer, Dordrecht
- Bentham J (1789) Introduction into the principles of morals and legislation. T. Payne, London, Also in Bowring J (ed) (1838) The works of Jeremy Bentham. W. Tait, Edinburgh
- Berger JO (1980) Statistical decision theory. Foundations, concepts, and methods. Springer, New York

- Bermúdez JL (2009) Decision theory and rationality. Oxford University Press, Oxford
- Bernoulli D (1738) Specimen Theoriae Novae de Mensura Sortis. Commentarii Academiae Scientiarum Imperialis Petropolitanae V [Papers of the Imperial Academy of Sciences in Petersburg, vol V]:175–192. Translated as: Bernoulli D (1954) Exposition of a new theory on the measurement of risk. *Econometrica* 22:23–36
- Binmore K (2009) Rational decisions. Princeton University Press, Princeton
- Blackwell D, Dubins L (1962) Merging of opinions with increasing information. *Annals of Statistical Mathematics* 33:882–886
- Bolker ED (1966) Functions resembling quotients of measures. *Trans Amer Math Soc* 124:292–312
- Bolker ED (1967) A simultaneous axiomatization of utility and subjective probability. *Philos Sci* 34:333–340
- Bradley R (1998) A representation theorem for a decision theory with conditionals. *Synthese* 116:187–229
- Bradley R (2001) Ramsey and the measurement of belief. In: Corfield D, Williamson J (eds) Foundations of Bayesianism. Kluwer, Dordrecht, pp 263–290
- Bratman M (1987) Intention, plans, and practical reason. Harvard University Press, Cambridge, MA
- Broome J (2002) Practical reasoning. In: Bermúdez J, Millar A (eds) Reason and nature: essays in the theory of rationality. Clarendon, Oxford, pp 85–111
- Camerer C (1995) Individual decision making. In: Kagel J, Roth AE (eds) Handbook of experimental economics. Princeton University Press, Princeton, pp 587–703
- Darwall S (ed) (2002) Consequentialism. With an introduction. Blackwell, Oxford
- Davidson D, Suppes P (1956) A finitistic axiomatization of subjective probability and utility. *Econometrica* 24:264–275
- Davidson D (1963) Actions, reasons and causes. *J Philos* 60:685–700
- De Finetti B (1931) Probabilismo. *Logos* 14:163–219. Translated as: (1989) Probabilism: a critical essay on the theory of probability and on the value of science. *Erkenntnis* 31:169–223
- De Finetti B (1937) La Prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré* 7:1–68. Translation as (1964) Foresight: its logical laws, its subjective sources. In: Kyburg HE, Smokler HE (eds) Studies in subjective probability, Wiley, New York, pp 53–118
- Dreier J (2004) Decision theory and morality. In: Mele A, Rawling P (eds) Oxford handbook of rationality. Oxford University Press, New York, pp 156–181
- Earman J (1992) Bayes or bust. MIT Press, Cambridge, MA
- Eells E (1982) Rational decision and causality. Cambridge University Press, Cambridge
- Ekenberg L, Boman M, Linnerooth-Bayer J (2001) General risk constraints. *J Risk Res* 4:31–47
- Ellsberg D (1961) Risk, ambiguity, and the savage axioms. *Q J Econ* 75:643–669
- Fehige C, Wessels U (1998) Preferences. De Gruyter, Berlin
- Fine T (1973) Theories of probability. An examination of foundations. Academic, New York/London
- Fishburn PC (1979) Utility theory for decision making. Krieger, Huntington, First edition, 1970
- Fishburn PC (1982) The foundations of expected utility. Reidel, Dordrecht
- Fishburn PC (1989) Retrospective on the utility theory of von Neumann and Morgenstern. *J Risk Uncertainty* 2:127–158
- Fishburn PC, Wakker P (1995) The invention of the independence condition for preferences. *Manag Sci* 41:1130–1144
- Foley R (1992) The Epistemology of belief and the epistemology of degrees of belief. *Am Philos Q* 29:111–121
- French S, Ríos Insua D (2000) Statistical decision theory. Arnold, London
- Friedman M, Savage LJ (1948) The utility analysis of choices involving risk. *J Political Econ* 56:279–304
- Friedman M, Savage LJ (1952) The expected utility hypothesis and the measurability of utility. *J Political Econ* 60:463–474
- Gardiner SM (2006) A core precautionary principle. *J Political Philos* 14:33–60
- Gibbard A, Harper W (1981) Counterfactuals and two kinds of expected utility. In: Harper W, Stalnaker R, Pearce G (eds) Ifs: conditionals, belief, decision, chance, and time. Reidel, Dordrecht, pp 153–190
- Gillies D (2000) Philosophical theories of probability. Routledge, London/New York
- Glover J (1990) Utilitarianism and its critics. Macmillan, New York
- Griffin J (1989) Well-being. Its meaning, measurement, and moral importance. Clarendon, Oxford
- Hacking I (1975) The emergence of modern probability. A philosophical study of early ideas about probability, induction and statistical inference. Cambridge University Press, London
- Hacking I (1990) The taming of chance. Cambridge University Press, Cambridge
- Hacking I (2001) An introduction to probability and inductive logic. Cambridge University Press, Cambridge
- Hájek A (2010) Interpretations of probability. In: Zalta EN (ed) The Stanford encyclopedia of philosophy (Spring 2010 Edition). <http://plato.stanford.edu/archives/spr2010/entries/probability-interpret/>
- Hammond PJ (1998a) Objective expected utility. In: Barberà S et al (eds) Handbook of utility theory, vol 1. Kluwer, Boston, pp 143–212

- Hammond PJ (1998b) Subjective expected utility. In: Barberà S et al (eds) *Handbook of utility theory*, vol I. Kluwer, Dordrecht, pp 213–271
- Hansson SO (2011) Risk. In: Zalta EN (ed) *The Stanford encyclopedia of philosophy* (Fall 2011 Edition). <http://plato.stanford.edu/archives/fall2011/entries/risk/>
- Hausner M (1954) Multidimensional utilities. In: Thrall RM, Coombs CH, Davis RL (eds) *Decision processes*. Wiley, New York
- Hawthorne J, Bovens L (1999) The preface, the lottery and the logic of belief. *Mind* 108:241–264
- Herstein IN, Milnor J (1953) An axiomatic approach to measuring utility. *Econometrica* 21:291–297
- Howson C (1995) Theories of probability. *Br J Philos Sci* 46:1–32
- Howson C, Urbach P (2006) *Scientific reasoning: the Bayesian approach*, 3rd edn. Open Court, La Salle
- Jaynes ET (1957) Information theory and statistical mechanics. *Phys Rev* 106:620–630
- Jaynes ET (1968) Prior probabilities. *IEEE Trans Syst Sci Cybern* 4/3:227–241
- Jaynes ET (1979) Where do we stand on maximum entropy? In: Levine RD, Tribus M (eds) *The maximum entropy formalism*. MIT Press, Cambridge, MA, pp 15–118
- Jeffrey RC (1965a) *The logic of decision*. University of Chicago Press, Chicago
- Jeffrey RC (1965b) New foundations for Bayesian decision theory. In: Bar-Hillel Y (ed) *Logic, methodology, and philosophy of science*. North Holland, Amsterdam, pp 289–300
- Jeffrey RC (2004) *Subjective probability: the real thing*. Cambridge University Press, Cambridge
- Jensen NE (1967) An introduction to Bernoullian utility theory. I. Utility functions. *Swed J Econ* 69:163–183
- Joyce J (1999) *The foundations of causal decision theory*. Cambridge University Press, Cambridge
- Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. *Econometrica* 47:263–291
- Kaplan M (1996) *Decision theory as philosophy*. Cambridge University Press, New York
- Kavka GS (1980) Deterrence, utility, and rational choice. *Theory and Decision* 12:41–60
- Keynes JM (1921) *A treatise on probability*. Macmillan, London
- Knight FH (1921) Risk, uncertainty and profit. Hard, Schaffner and Marx, Boston
- Kordsgaard CM (1997) The normativity of instrumental reason. In: Cullity G, Gaut B (eds) *Ethics and practical reason*. Clarendon, Oxford, pp 215–254
- Kreps DM (1988) Notes on the theory of choice. Westview Press, Boulder/London
- Kyburg HE, Smokler HE (eds) (1964) *Studies in subjective probability*. Wiley, New York
- Lewis D (1981) Causal decision theory. *Australas J Philos* 59:5–30
- Lewis D (1988) Desire as belief. *Mind* 97:323–332
- Lewis D (1996) Desire as belief II. *Mind* 105:303–313
- Lewis D (1997) Why conditionalize? In: *Papers in metaphysics and epistemology*. Cambridge University Press, Cambridge, pp 403–407
- Loomes G, Sugden R (1982) Regret theory: an alternative theory of rational choice under uncertainty. *Econ J* 92:805–824
- Luce RD, Raiffa H (1957) *Games and decisions: introduction and critical survey*. Wiley, New York
- Lumer C (2002) *The greenhouse. A welfare assessment and some morals*. University Press of America, Lanham
- Lumer C (2010) Rational choice and ethics. *Ethical Theory and Moral Practice* 13(5):483–593 (special issue with contributions by Lumer C, Narveson J, McClenen EF, Verbeek B, Hansson SO)
- Machina MJ (1982) “Expected utility” analysis without the independence axiom. *Econometrica* 50:277–323
- Makinson DC (1965) Paradox of the preface. *Analysis* 25:205–207
- Marschak JA (1950) Rational behavior, uncertain prospects, and measurable utility. *Econometrica* 18: 111–141
- McClenen E (2001) Bayesianism and independence. In: Corfield D, Williamson J (eds) *Foundations of Bayesianism*. Kluwer, Dordrecht, pp 291–307
- Mellor H (2005) *Probability. A philosophical introduction*. Routledge, London/New York
- Menger K (1931) Das Unsicherheitsmoment in der Wertlehre. *Betrachtungen im Anschluß an das sogenannte Petersburger Spiel*. *J Econ* 5:459–485
- Mill JS (1863) *Utilitarianism*. Parker, Son, and Bourn, London
- Nagel T (1970) *The possibility of altruism*. Oxford University Press, Oxford
- Peterson M (2008) Non-Bayesian decision theory. Beliefs and desires as reasons for actions. Springer, Berlin
- Peterson M (2009) *An introduction to decision theory*. Cambridge University Press, Cambridge
- Pfanzagl J (1967) Subjective probability derived from the Morgenstern-von Neumann utility theory. In: Shubik M (ed) *Essays in mathematical economics in honor of Oskar Morgenstern*. Princeton University Press, Princeton, pp 237–251
- Pratt JW (1964) Risk aversion in the small and in the large. *Econometrica* 32:122–136
- Price H (1989) Defending desire-as-belief. *Mind* 389:119–127
- Raiffa H (1968) *Decision analysis*. Addison-Wesley, Reading
- Ramsey FP (1931) Truth and probability. In: Braithwaite RB (ed) *Foundations of mathematics and other essays*. Routledge/P Kegan, London, pp 156–198. Reprinted in: Ramsey FP (1990) In: Mellor DH (ed).

- Philosophical papers. Cambridge University Press, Cambridge, pp 52–94
- Rawls J (1971) A theory of justice. Harvard University Press, Cambridge, MA
- Raz J (2005) The myth of instrumental rationality. *J Ethics Social Philos* 1:1 ([www.jesp.org](http://www.jesp.org))
- Resnik MD (1987) Choices. An introduction to decision theory. University of Minnesota Press, Minneapolis
- Robert CP (2007) The Bayesian choice. From decision-theoretic foundations to computational implementation. Springer, New York
- Samuelson PA, Nordhaus WD (2001) Economics, 17th edn. McGraw-Hill, Boston
- Savage LJ (1954) The foundations of statistics. Wiley, New York. Here quoted from 2nd edn. Dover, New York, 1972
- Scanlon TM (1998) What we owe to each other. Harvard University Press, Cambridge, MA
- Sen A, Williams B (1982) Utilitarianism and beyond. Cambridge University Press, Cambridge
- Shaw WH (1999) Contemporary ethics. Taking account of utilitarianism. Blackwell, Oxford
- Skyrms B (1999) Choice and chance: an introduction to inductive logic, 4th edn. Wadsworth, Belmont
- Smith M (1987) The Humean theory of motivation. *Mind* 96:36–61
- Smith M (2004) Instrumental desires, instrumental rationality. *Proceedings of the Aristotelian Society (Supplement)* 78:93–109
- Sobel JH (1994) Taking chances. Essays on rational choice. Cambridge University Press, New York
- Sugden R (1985) Why be consistent? A critical analysis of consistency requirements in choice theory. *Economica* 52:167–184
- Sugden R (2004) Alternatives to expected utility: foundations. In: Barberà S et al (eds) *Handbook of utility theory*, vol 2. Kluwer, Dordrecht, pp 685–755
- Suppes P (1956) The role of subjective probability and utility in decision-making. In: Proceedings of the third Berkeley symposium on mathematical statistics and probability, 1954–1955, vol V. University of California Press, Berkeley, pp 61–73
- Teller P (1973) Conditionalization and observation. *Synthese* 26:218–258
- Tversky A, Kahneman D (1981) The framing of decisions and the psychology of choice. *Science* 211:453–458
- von Neumann J, Morgenstern O (1947) Theory of games and economic behavior, 2nd edn. Princeton University Press, Princeton
- Wallace RJ (2001) Normativity, commitment, and instrumental reason. *Philosopher's Imprint* 1:1–26
- Weirich P (2004) Realistic decision theory: rules for nonideal agents in nonideal circumstances. Oxford University Press, New York
- Williamson J (2009) Philosophies of probability. In: Irvine A (ed) *Handbook of the philosophy of mathematics*. North Holland, Amsterdam, pp 493–533
- Williamson J (2010) In defence of objective Bayesianism. Oxford University Press, Oxford

# 16 A Philosophical Assessment of Decision Theory

Karsten Klint Jensen

Institute of Food and Resource Economics, University of Copenhagen,  
Frederiksberg C, Denmark

<b>Introduction .....</b>	<b>406</b>
<b>A Short History of Decision Theory and the Notion of Utility .....</b>	<b>408</b>
Bernoulli's Hypothesis .....	408
Bentham and the Utilitarian Economists .....	410
The Axiomatic Turn .....	411
<i>Interlude: Measurement Theory .....</i>	<i>412</i>
Von Neumann–Morgenstern Utility .....	413
Early Modern Decision Theory .....	414
<b>Savage's Decision Theory .....</b>	<b>415</b>
The Basic Concepts .....	415
The Sure-Thing Principle .....	417
Qualitative and Quantitative Subjective Probability .....	419
The Representation and Uniqueness Theorems .....	420
The Maxim "Maximize Expected Utility" .....	421
The Problem of Action Guidance .....	422
<b>Decision Theory as a Theory of Good .....</b>	<b>424</b>
External Reasons .....	424
Probability-Relative Goodness .....	426
Problems for the Maxim of Maximizing Expected Goodness .....	427
Bernoulli's Hypothesis .....	428
<b>General Good and Individual Good .....</b>	<b>429</b>
Equalizing Risk .....	430
The Principle of Personal Good .....	431
Valuing Life .....	433
The Interpersonal Addition Theorem .....	434
Equality in Outcomes .....	436
<b>Further Research .....</b>	<b>437</b>

**Abstract:** The significance of decision theory consists of giving an account of rational decision making under circumstances of uncertainty. This question is important both from the point of view of what is in our personal interest and from the point of view of what is ethically right. But decision theory is often poorly understood and its significance only sparsely discussed in the literature.

In a short history of decision theory, it is demonstrated how modern axiomatic decision theory works differently from classical decision theory, but also how it is confused with it. Further, it is explained how modern axiomatic decision theory is an instance of fundamental measurement theory. This is then followed by a thorough introduction to Savage's version of modern axiomatic decision theory.

Turning to the interpretation of the theory, the maxim "maximize expected utility," which stems from classical decision theory, is shown to misrepresent the structure of modern axiomatic decision theory. Whereas the classical theory *assumes a value assignment to outcomes and derives preferences over uncertain acts*, the modern axiomatic approach *assumes preferences over uncertain acts and derives the utility assignments*. In the modern approach, the action guidance is to conform to the axioms.

Analyzing decision theory as a theory of good, the maxim "maximize expected goodness" repeats the misunderstanding. Moreover, it implies risk neutrality about good and a cardinal measure of good, and both are problematic. Only an ordering of uncertain acts that conforms to the axioms allows for risk aversion about good. If there were an independent cardinal measure of the goodness of outcomes, utility would be an increasing, strictly concave transform of good.

The Principle of Personal Good states the idea that the ordering of uncertain acts according to general betterness should be determined by how good the uncertain acts are for individuals. It sounds like a reasonable idea, and a widely used way of valuing life is based on it. But it is certainly not uncontroversial because it conceals conflicts of interest between individuals in final outcomes. In the context of decision theory for general and individual betterness, the Principle of Personal Good holds if, and only if, general utility is the sum of individual utilities.

The chapter concludes with suggestions for future research.

---

## Introduction

The significance of decision theory consists of giving an account of rational decision making under circumstances of uncertainty. This is clearly an issue of great practical importance. Almost any decision one can think of is taken in a context in which the decision maker has only limited information. Therefore, the decision maker is faced with more or less uncertainty about what the actual outcome of the decision will be. This again means that he is faced with the risk that the outcome might be different from the one he intends, possibly with worse consequences. How should we deal with this uncertainty and the risks it involves? This question is important both from the point of view of what is in our personal interest and from the point of view of what is ethically right.

Decision theory is supposed to enable us to deal rationally with uncertainty. It tells us how to model the uncertainty of a decision situation and thereby how to structure a decision problem. It is, moreover, supposed to guide us by explaining how the rational decision maker ought to choose in the face of uncertainty (I will only be concerned with decision theory as

a normative theory). However, even though many are familiar with the maxim “maximize expected utility,” there is, perhaps surprisingly, great confusion about what the actual message of decision theory really is.

On the face of it the confusion stems from the fact that the term “utility” is used with many different meanings, which are often not made clear and, moreover, often confused with one another. The core of modern axiomatic decision theory is a pair of formal theorems. In the standard interpretation, they state that if the decision maker’s preferences fulfill certain axioms (requirements of rationality), then these preferences can be represented by an expected utility function, and this function is *cardinal* (I shall explain later what all this means). This meaning of “utility” (a cardinal function representing preferences) is often poorly understood by people not fully acquainted with the formal content of decision theory. And it is often used alongside and confused with “utility” referring to what is “good for an individual,” in one sense or another, or to other forms of substantial value.

Of course, the confusion of meanings is unhelpful. But there is also a substantial problem lying underneath this surface. As a pure formal theory, decision theory does not offer us any guidance. The formal apparatus needs interpretation in order to be of any help. But once we look for a plausible interpretation that could guide us prudentially as well as morally, the question of how the technical apparatus of decision theory should be combined with relevant theories of prudential and ethical value emerges. This again raises the question of how the formal notion of “utility” relates to substantial values.

Whereas the axioms of decision theory, and other aspects of the formalization, are a matter of some controversy, which has been discussed at length in the literature (much of which is described elsewhere in this book), the question about what can be learned from decision theory in the context of prudential and ethical value has, on the contrary, only been briefly addressed. However, this has profound implications for our understanding of how to deal with risk. Hence, in this chapter, I shall take decision theory and its axioms for granted and attempt to identify a plausible interpretation of the theory, as applied to what is good for an individual and what is generally good.

An additional source of confusion is that there are several versions of decision theory. I shall be concerned with what I call *axiomatic expected utility theory*. The notion of expected utility involves probabilities. Some versions of decision theory do not have the expected utility form, because they are concerned with situations in which no probabilities are available (a condition often called “ignorance”).

Among expected utility theories, there is a distinction between classical and modern theories. Classical theories established an objective (which later came to be called “utility”), which a rational individual ought to maximize. Modern theories work in a different way. They establish axioms about the structure of preferences and claim that a rational individual should have preferences that satisfy them. The theories show that if an individual does satisfy the axioms, then “utilities” can be constructed such that the individual maximizes his expected utility (I shall say much more on this difference below).

Different modern theories have different axioms. However, what I would like to call the conventional theories, are sufficiently similar to allow me to speak generally about “decision theory.” There are some unconventional theories as well, which deviate from the general picture (e.g., Jeffrey 1983), or even take a non-expectational form (e.g., Machina 1982), but I shall leave these aside. I shall use Leonard Savage’s theory (Savage 1972) as representative of “decision theory” in my discussion, because I consider it to be the most general statement of the ideas underlying conventional decision theory.

The chapter first gives a short account of the history of decision theory in order to locate the origins of the different meanings of “utility” and to explain the development from the classical theories to the modern axiomatic approach. Savage’s theory is presented in some detail, concentrating on the aspects of importance for rest of the chapter. The structure of the theory is explained and contrasted with a widespread misrepresentation.

Next, decision theory is examined as a theory of what is good for an individual and what is ethically good in the face of uncertainty. Concerning the latter, a specific discussion of how the ethical attitude toward uncertainty should reflect what is good for individuals in the face of uncertainty is presented. The chapter concludes by outlining questions for further research.

Decision theory is a formal theory, and many issues relating to it involve further formal complications. A few formal issues are important for the main argument of the chapter. The ideas behind these issues are explained such that they should be intuitively accessible, even for readers with a limited background in mathematics.

In some cases, I shall also mention formal issues related to decision theory, but I will not go into the details. Some readers will be acquainted with these details, while others will not. However, readers unfamiliar with the complexities can safely skip these instances of “name dropping” without causing problems for their understanding.

John Broome’s writings play a prominent role in this chapter, because he, more than anyone else, has discussed the philosophical aspects of decision theory. I am very grateful to the Swedish Science Council who supported part of the research underlying this chapter.

---

## A Short History of Decision Theory and the Notion of Utility

---

### Bernoulli’s Hypothesis

---

The first developments of relevance for this chapter date back to the seventeenth century when the theory of probability was founded. These early theorists were largely concerned with the question of how to evaluate gambles of money. They concluded with the maxim that the gamble with the greatest expected winnings was the most advantageous. This maxim, known as the *Principle of Expected Value* or the *Principle of Mathematical Expectation*, states that we should evaluate a gamble according to the sum of its possible winnings or losses, each weighted with its probability.

The Principle of Expected Value involves risk neutrality about money. *Risk neutrality about money* means that if two gambles have the same expected monetary value, then they are always equally good. But in some cases, it seems reasonable to prefer avoiding the risk of losing. *Risk aversion about money* means that if two gambles have the same expected monetary value, the less risky is always better. Daniel Bernoulli (1738) was the first to question the Principle of Expected Value (translated in Bernoulli (1954), quoted from Page (1968), p. 200):

- ▶ Somehow a very poor fellow obtains a lottery ticket that will yield with equal probability either nothing or twenty thousand ducats. Will this man evaluate his chance of winning at ten thousand ducats? he asks. Would he not be ill-advised to sell this lottery ticket for nine thousand ducats?

Bernoulli thinks the answer is “no.” On the other hand, a rich man might be more prone to risk. Bernoulli draws the conclusion that the value of an item to an individual should not be

based on its monetary price. Rather, it should be based on what Bernoulli calls *emolumentum*. The *emolumentum* of a gamble is not the same for all individuals; it depends on the individual's wealth. A gain of 1,000 ducats is more significant to the poor man than it is to the rich man.

In the translation of Bernoulli's paper in *Econometrica* from 1954, *emolumentum* was translated as "utility." This translation may be one of the sources of the confusion of Bernoulli's theory with modern axiomatic utility theory. *Emolumentum* means something like *advantage* or *benefit*. The value to an individual of an item is thus the degree to which it is in his interest, or how good it is for him. Bernoulli suggests the maxim that an individual should evaluate a gamble by its *emolumentum medium*, i.e., its expected goodness to him. John Broome (1991b) has dubbed this maxim *Bernoulli's Hypothesis*.

Gabriel Cramer, whom Bernoulli quotes in French in his paper and who appears to be the one who first presented these ideas, used the expression *Esperance Morale* (moral expectation) for what Bernoulli calls *emolumentum medium*. Because of this, the translation "moral worth" is sometimes used instead of the translation "utility." Cramer stressed the moral character of the question of how to measure the true value of an item to an individual. He further identified the moral expectation of a certain gain with the pleasure one hopes to derive from it and, correspondingly, the moral expectation of a loss with the pain caused by it.

To summarize Bernoulli's theory: In the face of risk, an individual should be concerned with his own good. Money has diminishing marginal goodness to an individual. Therefore, an individual has reason to be risk averse concerning money; the poorer he is, the more he has reason to prefer money for sure to gambles with a risk of losing. To this, Cramer adds the idea that the goodness of an item for an individual is the pleasure one can derive from it.

Bernoulli used his theory to solve the so-called St. Petersburg Paradox, which had been presented to him by his cousin Nicolas Bernoulli (quoted from Page (1968), p. 209):

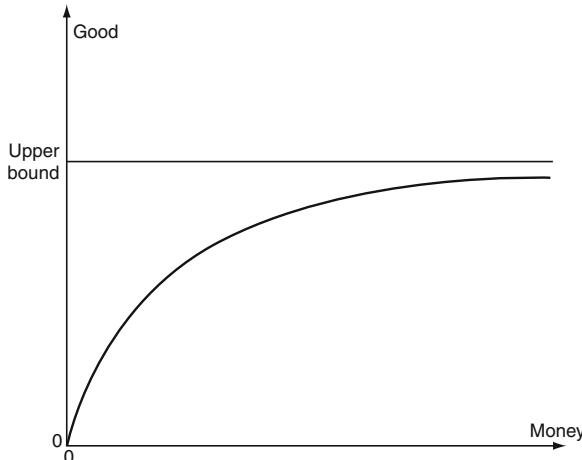
- Peter tosses a coin and continues to do so until it should land "heads" when it comes to the ground. He agrees to give Paul one ducat if he gets "heads" on the very first throw, two ducats if he gets it on the second, four ducats if on the third, eight if on the fourth, and so on, so that with each additional throw the number of ducats he must pay is doubled.

How should we evaluate this gamble? Notice that the expected value of the gamble is infinitely high:

$$\frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 4 + \dots = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \dots = \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n \cdot 2^{n-1} \rightarrow \infty$$

The "paradox" is that according to the Principle of Expected Value, Paul should be willing to pay any finite sum for this gamble. However, no one would pay more than a small price (Bernoulli suggests 20 ducats). The solution hinted at by Bernoulli, and more clearly stated by Cramer, consists of the claim that money has diminishing marginal goodness or moral worth to an individual, which moreover has an upper bound when one's wealth grows toward infinity. This solution has been widely accepted (e.g., Savage (1972) and Arrow (1970)), although whether it really answers the problem of the paradoxically high expected value of this kind of gamble, is still contested. However, we shall not pursue this question any further.

The diminishing marginal goodness of money implies that goodness as a function of money is an increasing strictly concave function. The graph of an increasing strictly concave function has a characteristic downward curvature. Cramer further assumed that the function has an upper bound, see Fig. 16.1.



**Fig. 16.1**  
The goodness of money

### Bentham and the Utilitarian Economists

Economists adopted the notion of “utility” for choices under certainty. Savage (1972, p. 95) suggests that Bernoulli’s paper might have been the principal source for the introduction of the notion of “utility” in economics. At any rate, the utilitarian philosopher Jeremy Bentham (1789) connected the classical economists’ use of “utility” with the neoclassical economists’ use by saying that (p. 2):

- ▶ By utility is meant that property in any object, whereby it tends to produce benefit, advantage, pleasure, good, or happiness, (all this in the present case comes to the same thing) or (what comes again to the same thing) to prevent the happening of mischief, pain, evil, or unhappiness to the party whose interest is considered.

In other words, an object has utility to the extent that it promotes what is in an individual’s interest, or what is good for the individual. Bentham goes on to determine what is good for an individual more specifically as, “pleasure” (philosophers call this account of an individual’s good *hedonism*). A pleasure can be measured (at least in theory) by its intensity and duration, and amounts of pleasure can be added up and compared across people. Bentham’s theory was adapted by W. S. Jevons in his *Theory of Political Economy* (1873). Commenting on Bentham’s definition of “utility,” Jevons (1873/1970, p. 46) said: “This perfectly expresses the meaning of the term in Economy”.

After this, the meaning of “utility” in economics gradually changed from the tendency of an object to produce good to the good itself. An individual’s “utility” is an individual’s own good. This change is documented in Broome (1991a), reprinted in Broome (1999), pp. 19–28. Since an individual’s good was identified with “pleasure,” utility was considered a mental entity, the amount of which could be measured and compared across people.

A *rational* individual should prefer the outcome that provides him with the highest utility, i.e., the greatest (expected) balance of pleasure over pain. The “law of diminishing marginal

utility” was widely accepted, not just for money as in Bernoulli’s case, but for all goods. As a consequence, economists claimed that a transfer of income from a rich person to a poor person would increase the total utility of society, other things being equal, and therefore represent a better overall outcome.

## The Axiomatic Turn

---

During the 1920s, economists became increasingly skeptical about the notion of utility as a mental entity which could be measured. For one thing, the measurement of mental entities was never made operational. But perhaps more important was the dawning insight into utility functions. An individual’s utility function was supposed to measure the balance of pleasure over pain for the various available alternatives and if the individual was rational, he should form his preferences accordingly. However, it was discovered that for a function representing an individual’s preferences over outcomes, only the *ordering* of the outcomes is significant. Any function preserving the same order represents these preferences equally well. This is called an *ordinal* utility function.

Since the basic economic theorems could be proved on the basis of ordinal utility functions alone, the notion of utility as a measurable mental entity became theoretically superfluous for explaining an individual’s preferences. The preferences could simply be considered basic in the theory. This development is complete in Hicks and Allen (1934).

Moreover, there was also skepticism concerning the comparison of utility as pleasure across individuals. Many economists considered such comparisons meaningless, because they did not believe there was any fact of the matter. Or if such comparisons were accepted as meaningful, they were deemed *value judgments* that belonged to the realm of ethics and not the realm of economic science – again because they appeared not to be factual statements (see, e.g., the discussion in Robbins (1935)). This explains the prominence of the notion of a *Pareto Improvement* in economics ever since, which only applies when everyone prefers one state to another or is indifferent between them, and at least one individual definitely prefers it. Such comparisons of states are not dependent on interpersonal comparisons.

Through this development, “utility” came to mean the value of a function representing an individual’s preferences. “Represents” in this sense simply means that one outcome has a higher utility than another if, and only if, it is preferred to it. Consider the weak preference relation “– is weakly preferred to –” (i.e. “– is preferred or indifferent to –”). A utility function for the case of a finite, or infinite but countable, set of outcomes exists if, and only if, an individual’s weak preference relation for outcomes are transitive and complete. The proof of this dates back to Cantor (1895). For an uncountable set of outcomes, a further continuity condition is needed Debreu (1959).

A weak preference relation is *transitive* if, and only if, for all triples of outcomes  $A$ ,  $B$ , and  $C$ , if  $A$  is weakly preferred to  $B$ , and  $B$  is weakly preferred to  $C$ , then  $A$  is weakly preferred to  $C$ . Transitivity is widely considered a rational requirement of consistency.

A weak preference relation is *complete* if, and only if, for all pairs of outcomes  $A$  and  $B$ , either  $A$  is weakly preferred or indifferent to  $B$ , or  $B$  is weakly preferred or indifferent to  $A$ . Completeness implies that all outcomes are comparable; either one outcome is preferred to another, or the other way round, or there is indifference between them – no other possibilities are allowed.

It is implicit in this change that a *rational* individual is now defined as a person whose preferences are transitive and complete. Whereas before, a rational individual's utility function (measuring the total pleasure) was given and his preferences derived from it, preferences are now given as primitives, from which the utility function (now understood as numbers representing these preferences) is derived.

The modern definition of "utility" is thus *the value of a function which represents a person's preferences*. This is a purely technical definition of "utility"; there is no logical connection to an individual's good. However, economists very often assume a connection of this sort. That is, they assume *The Preference Satisfaction Theory* about an individual's good: An outcome is better for an individual than another if, and only if, it is preferred. This theory implies that an outcome is better for an individual if, and only if, it has a higher utility. Thereby, the new meaning of utility is confused with the old meaning, which refers to an individual's good.

The Preference Satisfaction Theory only mentions an individual's preferences. Either the preferred outcome obtains or it does not. No reference is made to any mental entity such as "pleasure." However, the Preference Satisfaction Theory about an individual's good is often confused with the old hedonistic theory about an individual's good when the discussion uses expressions like *experienced* or *felt satisfaction* (i.e., a mental state) of the preferred outcome. The confusion about "utility" is thereby doubled.

### **Interlude: Measurement Theory**

---

The formal content of the axiomatic utility theory belongs to fundamental measurement theory (Krantz et al. 1971), which investigates the conditions under which empirical comparisons of properties of objects can be faithfully transformed into measurement scales. Examples of empirical relations are "longer than," "hotter than," or as in our case, "is weakly preferred to."

Given an empirical relation on a set of objects, a *representation theorem* asserts that *if* the empirical relation satisfies certain axioms, *then* a *numerical representation* can be constructed, i.e., a real-valued function that assigns numbers to the objects in such a way that the size of the numbers preserves the empirical ordering of the objects. In our case, a numerical representation can be constructed, if the weak preference relation satisfies the axioms of transitivity and completeness.

A *uniqueness theorem* states what is unique about this numerical representation. It says that *if* a certain function is a representation of the empirical relation in question, *then* another function represents the same empirical relation if, and only if, it is a certain transformation of the first. Hence, the uniqueness of a numerical representation is determined by all and only the kind of transformations that preserve the representation of the empirical relation. The uniqueness of the representation, in turn, defines the kind of measurement scale involved in the representation.

For an *ordinal scale*, only the order of the numbers is unique, anything but the order is arbitrary. An ordinal scale is *unique up to increasing transformation*, i.e., a function represents the same empirical structure as another if, and only if, it is an increasing transform of it. An increasing transformation  $f$  of a function  $u(x)$  has the form  $f(u(x))$ , where  $f$  is an increasing function. An increasing transformation can change everything except the order of the numbers (an example would be if  $u(x) = 1$ ,  $u(y) = 2$ ,  $u(z) = 3$  were transformed to  $f(u(x)) = 10$ ,  $f(u(y)) = 100$ ,  $f(u(z)) = 101$ ).

For an *interval scale*, the choice of unit and zero is arbitrary, but ratios of intervals are uniquely determined. An interval scale is *unique up to increasing linear transformation*, i.e., a function represents the same empirical structure as another if, and only if, it is an increasing linear transform of it. An increasing linear transform of a function  $u(x)$  has the form  $au(x)+b$ , in which  $a$  is a positive number. A linear transform implies that the zero and the unit are changed, but the ratio between intervals remains constant (an example would be switching from the Fahrenheit to the Celsius temperature scale), so:

$${}^{\circ}\text{C} = \frac{5}{9}({}^{\circ}\text{F} - 32).$$

Finally, for a *ratio scale*, the zero is also uniquely determined. A ratio scale is *unique up to similarity transformation*, i.e., a function represents the same empirical structure as another if, and only if, it is a similarity transform of it. A similarity transformation of a function  $u(x)$  has the form  $au(x)$ , where  $a$  is a positive number. A similarity transform implies that the unit is changed, but ratios of numbers remain constant (an example would be switching from centimeters to inches).

In our case, the utility function representing the weak preference relation is an ordinal scale. Interval and ratio scales are both called *cardinal scales*. We shall see examples of cardinal utility scales below.

## Von Neumann–Morgenstern Utility

---

In 1944, John von Neumann and Oskar Morgenstern published the seminal *Theory of Games and Economic Behaviour*. Roughly, games are situations of mutual interdependence, in which the outcome of an individual's choice is dependent on how the others choose to act. Each player is assumed to have a payoff function, defined on all possible outcomes of the game; and each player is further assumed to have full knowledge of the payoff functions of all players. Thus, there is no doubt about what the players prefer in terms of final outcomes. However, the best strategy for a player remains to be determined, because there is uncertainty about the other players' moves. Each player knows that the others will try to maximize their payoff, and they have to consider the likely moves of others when choosing their own. The task for game theory is thus to find the best strategy for players in different types of games.

I shall not be concerned with game theory per se. However, as a side issue, von Neumann and Morgenstern made an important contribution to decision theory. In their game theory, they assumed that a payoff function was based on the players' preferences over outcomes. But they also assumed that this function, which they also called a "utility function," not just preserved the preference ordering, but in fact – as they unfortunately described it – was an instance of "measurable utility." To justify this assumption, they presented a new form of axiomatic utility theory (the proof of which they did not give until the second edition in 1947).

The axiomatic utility theory in economics discussed above belongs to what we might call decision making under certainty. In this case, there is no distinction between an act and the outcome it leads to. Von Neumann and Morgenstern considered preferences over *risky acts*. A risky act is one in which the decision maker does not know the outcome with certainty. He knows that the act has a number of possible outcomes, each of which can occur with a known probability. Such risky acts are often called *lotteries*.

As in the case of decision making under certainty, von Neumann and Morgenstern wanted to represent the decision maker's preferences by a function. But they did not just require that an act had a higher utility if, and only if, it was preferred, they also required that the utility  $u$  of an act, which gave outcome  $A$  with probability  $p$  and outcome  $B$  with probability  $(1-p)$ , should be expressed as  $pu(A)+(1-p)u(B)$ , i.e., as the expected utility of the act.

In other words, the domain on which the utility function was defined had a richer structure than in the case of decision making under certainty. It contained the operation of adding up the utilities of each possible outcome, weighted by their probabilities, to give the utility of the risky act. Von Neumann and Morgenstern could then prove that if a utility function, which satisfies both these conditions exists then it is determined up to linear transformation; i.e., another utility function only represents the same preferences by fulfilling the two requirements if, and only if, it is a linear transformation of the first function. And this makes utility a “measurable entity,” not in the sense of measuring a mental entity such as pleasure, but in the sense of a cardinal utility function that represents preferences.

In an appendix to the second edition, von Neumann and Morgenstern then showed the axioms for preferences over risky acts that are necessary and sufficient to prove the existence of an expected utility function. I shall not give an account of the axioms here. Instead, I will present the axioms of Savage's more general decision theory below.

There has been much confusion about von Neumann and Morgenstern's utility theory. It is clearly an axiomatic utility theory, which states that if given preferences over risky acts satisfy certain axioms, they can be represented by an expected utility function. However, their reference to “measurable utility” made many economists believe that they had attempted to reintroduce utility as a measurable mental entity, which the decision maker aims to maximize.

Savage (1972), who presented the axiomatic decision theory I shall be turning to later on, added to the confusion by stating that von Neumann and Morgenstern gave “strong intuitive grounds for accepting the Bernoullian utility hypothesis as a consequence of well-accepted maxims of behavior” (p. 97). However, Bernoulli meant “an individual's good” by “utility,” whereas von Neumann and Morgenstern meant the value of a function representing preferences over risky acts. In fact, somewhat ironically, von Neumann and Morgenstern (1944/2004, p. 8) assumed that agents maximize their monetary outcome, which they – contrary to Bernoulli – considered “identical, even in the quantitative sense, with whatever ‘satisfaction’ or ‘utility’ is desired by each participant”; in other words, they assumed risk neutrality about money. Clearly, this did not make use of the full potential of their decision theory and – as we shall see below – probably added to the widespread confusion in the area. (Note also that von Neumann and Morgenstern spoke of “desired utility,” which they identified with “desired satisfaction,” another indication of the high degree of confusion in this area.)

---

## Early Modern Decision Theory

Von Neumann and Morgenstern's decision theory assumed preferences over risky acts as basic. But in game theory, they wanted to calculate preferences over acts, given the players' payoff functions for outcomes. And the next developments in decision theory were modeled on game theory. Thus, in a few places (e.g., pp. 32, 85–86), von Neumann and Morgenstern (1944/2004) hint at the idea of a *one person statistical game*, in which the decision maker aims to maximize his payback function in a choice between a number of risky acts.

Abraham Wald adapted this idea in his *Statistical Decision Functions* (1950). Wald's idea was that the concept of game against "nature" (a "player" who does not aim to maximize "his" outcome, but who "chooses" according to a certain probability distribution for outcomes) could serve as a new foundation for statistics. In statistics, the testing of hypotheses and point estimation are treated as separate issues. Wald suggested that they could be seen as instances of the more general problem of choosing a good decision strategy. But note that Wald assumed that the decision maker (the statistician) had a payoff function, which was defined on outcomes.

Von Neumann and Morgenstern's decision theory is often described as being concerned with decision making under risk, because it assumed known probabilities for each possible outcome. In the 1950s, decision theory for decisions under uncertainty was founded. Uncertainty was understood as a situation in which probabilities for "nature's choices" (often called *states of nature*) are unknown, and perhaps do not even exist. Again, game theory was the model. The decision maker has complete preferences over outcomes and he looks for the strategy (i.e., preferences over uncertain *acts*), which is likely to best satisfy these preferences. The first strategy to be examined was the *minimax strategy* (an agent chooses the act so that maximal loss – from the worst possible outcome – is minimized) – a strategy which was in fact adapted from the study of zero-sum games in game theory.

However, this is not the form of decision theory I shall be concerned with here. In 1954, Leonard Savage presented his *The Foundations of Statistics*, which was another form of decision theory for decisions under uncertainty. Savage built on Wald's idea of statistical decision theory, but he took preferences over uncertain acts to be basic. In contrast to the earlier decision theory for decisions under uncertainty, he added the idea of subjective probabilities (and his theory is often called *Bayesian decision theory*). This means that in situations where objective probabilities are unknown or do not exist, the decision maker can still (and according to the theory should) form beliefs about the likelihoods of the various states of nature. Insofar as a consistent measure for subjective probabilities can be construed, decision making under uncertainty can be reduced to a form similar to decision making under risk.

Similar ideas had actually been put forward by Frank Ramsey as early as 1926, but his paper was not published until 1931 after his death, and it did not have much impact at that time. Savage's theory is more general. His account of subjective probabilities was largely inspired by a paper by Bruno de Finetti (1937), translated in de Finetti (1964).

Decision theorists who believe that subjective probabilities are arbitrary in cases where there is no evidence whatsoever have continued to develop strategies in the tradition of the early theories of decision making under uncertainty. They now call these situations *ignorance*. For more on such theories see, e.g., Luce and Raiffa (1957, pp. 275–326). I shall now introduce what I shall use as the paradigm of modern axiomatic decision theory for the rest of this chapter.

## Savage's Decision Theory

### The Basic Concepts

Savage (1972) models uncertainty in the following way: The individual in a decision situation is concerned with some object (called the *world*) because the outcome of the acts, he has to choose between, is determined by the state of this world, but he does not know which of several

possible states is the true state. A *state* is thus a full description of the relevant aspect of the world resolving all uncertainty. The *true state* is the state that does in fact obtain. An *event* is a set of states.

An *act* is defined as a function that attaches an outcome to each state of the world. An *outcome* is a description of what finally matters to the decision maker, i.e., how his life and well-being are affected (Savage uses the term “consequence,” but I shall use the term “outcome”). An act is thus exclusively determined by its possible outcomes, it does not have any properties that do not show up in some outcome.

Savage provides this example to illustrate the concepts (pp. 13–14):

- ▶ Your wife has just broken five good eggs into a bowl when you come in and volunteer to finish making the omelet. A sixth egg, which for some reason must either be used for the omelet or wasted altogether, lies unbroken beside the bowl.

The uncertainty in this case is about whether the sixth egg is good or rotten. Savage suggests that you must decide among three acts, namely, to break it into the bowl with the other five, to break it into a saucer for inspection, or to throw it away without inspection. Depending on the state of the egg, each of these acts will have some consequence for you, as outlined in this decision matrix (❷ *Table 16.1*).

There are some important assumptions underlying this model of uncertainty. States are assumed to be mutually exclusive, and the specification of states is assumed to be exhaustive. This ensures that one and only one state obtains. Also, the acts are assumed to be mutually exclusive. And finally, states are assumed to be causally independent of acts. Richard Jeffrey (1983) suggested a decision theory without this latter restriction, which might therefore be considered more general. However, it also involves some complications which I do not want to go into here.

There are no further rules for the specification of a decision problem. The specification of a decision problem is relative to the description of states. A description of a state cannot cover all aspects. Hence, there will be further uncertainties under the surface of a description resolving the uncertainty in focus. Savage says that any small world state is an act in a larger world, which means that further uncertainties can be analyzed as the possible outcomes in a finer-grained description of states.

I shall now present the important details of Savage’s theory. Suppose there are  $n$  acts and  $s$  states of the world. Each act  $A$  assigns an outcome to each state  $(a_1, a_2, \dots, a_s)$ . The theory is

❷ **Table 16.1**

An example illustrating the basic concepts

States Acts	6th egg is good	6th egg is rotten
Break 6th egg into bowl	Six-egg omelet	No omelet, and five good eggs destroyed
Break 6th egg into saucer for inspection	Six-egg omelet, and a saucer to wash	Five-egg omelet, and a saucer to wash
Throw 6th egg away	Five-egg omelet, and one good egg destroyed	Five-egg omelet

divided into two parts. One part is concerned with the existence of a probability measure that assigns unique probabilities  $p_1, p_2, \dots, p_s$  to each state that sum to one. The other part is concerned with assigning utilities to outcomes such that, given the probabilities for states, the utility of an act  $A$  is its expected utility:

$$u(A) = p_1 u(a_1) + p_2 u(a_2) + \dots + p_s u(a_s)$$

Given the probability measure, this latter part is almost identical to von Neumann and Morgenstern's theory for decisions under risk.

Savage now goes on to assume that the decision maker has a weak preference relation defined on all pairs of acts. The *first axiom* states that the weak preference relation for acts is *transitive* and *complete*. We recognize this requirement from axiomatic decision theory for decisions under certainty described above. It ensures a complete ordering of all acts.

## The Sure-Thing Principle

---

Next, Savage defines the notion of one act being weakly preferred to another *given some event*. Remember that there are  $s$  states and that an event is a subset of states. Consider a subset of states  $X$ , e.g.,  $s_1$  and  $s_3$  (the event that either  $s_1$  or  $s_3$  obtains). Consider further two acts  $A$  and  $B$ , which have the same outcomes in the event that  $X$  does not obtain ( $\sim X$ ), i.e., in all other states  $s_2, s_4, s_5, \dots, s_s$ .  $A$  is weakly preferred to  $B$  given  $X$  if, and only if,  $A$  is weakly preferred to  $B$ .

The *second axiom* is called the *Sure-thing Principle*. Suppose we have four acts  $A, B, C$ , and  $D$ . For a fixed set of states,  $A$  and  $B$  have identical outcomes, but they differ in their outcomes for the remaining states. For the same fixed states,  $C$  and  $D$  likewise have identical outcomes, but not necessarily the same as  $A$  and  $B$ . For the rest of the states,  $C$  has outcomes that are identical with  $A$ 's in these states, and  $D$  has outcomes that are identical with  $B$ 's outcomes in these states. The Sure-thing Principle now states that if  $A$  is weakly preferred to  $B$ , then  $C$  is weakly preferred to  $D$ .

This can also be expressed thus: if one act is weakly preferred to another, given some event  $X$ , then this preference holds, whatever (identical) outcomes occur in the event  $\sim X$ . The essence of the Sure-thing principle is that, when comparing two acts given some set of states, the evaluation can be done independently of what happens in the remaining states. Thus, the reasons for preferring  $A$  to  $B$  must come from the set of states, where they differ in outcomes. The same can be said about  $C$  and  $D$ . And since the reasons for preferring  $A$  to  $B$  are the same as the reasons for preferring  $C$  to  $D$ , it follows that if an individual prefers  $A$  to  $B$ , he/she should also prefer  $C$  to  $D$ .

In order to see what this excludes, consider the famous counter-example given by Maurice Allais (1953), translated in Allais (1979). You have to decide your preferences among the following lotteries, in which one out of a hundred numbered tickets is drawn randomly, and you receive a prize determined by the number of the ticket (☞ [Table 16.2](#)).

Allais prefers  $A$  to  $B$  and so do many other people as several empirical studies have shown. Their reason is that  $A$  gives €1 million with certainty, whereas  $B$  involves a small risk of getting 0. Since  $A$  and  $B$  have identical outcomes for states 12–100, we can say that they prefer  $A$  to  $B$  given states 1–11. But  $C$  and  $D$  likewise have identical outcomes for states 12–100; and given states 1–11,  $A$  is identical with  $C$ , and  $B$  is identical with  $D$ ; therefore, the Sure-thing Principle

■ **Table 16.2**  
Allais' example

Lottery A			Lottery C		
States			States		
Ticket 1	Tickets 2–11	Tickets 12–100	Ticket 1	Tickets 2–11	Tickets 12–100
€1 million	€1 million	€1 million	€1 million	€1 million	0
Lottery B			Lottery D		
States			States		
Ticket 1	Tickets 2–11	Tickets 12–100	Ticket 1	Tickets 2–11	Tickets 12–100
0	€5 million	€1 million	€0 million	€5 million	0

implies that they should also prefer *C* to *D*. But Allais (and many others besides him) prefers *D* to *C*. For him, since there is an 89% chance of getting 0 anyway, the 10% chance of getting €5 million outweighs the extra 1% risk of getting 0. So this preference violates the Sure-thing Principle.

The Sure-thing Principle implies that what happens in one state can be evaluated separately from what happens in the others. There can be no interaction between states in the evaluation. A way to justify this principle is to point out that only one state can obtain. When we evaluate a state, we should imagine what it would be like if that state obtains. And if it does, no other state will occur. Therefore, how could what never happens possibly affect the value of what does happen?

But Allais suggested that there can be interactions: The reason to prefer *A* to *B* is that *A* gives €1 million *with certainty*, and certainty is something that does not show up in any state separately, it results from what happens in all states taken together.

Decision theorists feeling pressed by Allais' or similar examples do have a maneuver in response. If there is something special about certainty, it should show up as part of some outcome. For instance, it has been suggested that if you get 0 in *B*, you would feel very disappointed. But then this disappointment should be part of the outcome. And if the outcome in *B* for state ticket 1 is "0 and feeling very disappointed," there is no longer any threat to the Sure-thing Principle. The problem with this maneuver is that it should not be ad hoc. We would like to have some nonarbitrary, independent principles for when more fine-grained descriptions of outcomes are allowed.

It is not the task of this chapter to take a stand for or against the Sure-thing Principle. It is rather to explore the consequences of accepting the Sure-thing Principle. I shall just point out that the Sure-thing Principle and the idea of giving special weight to certainty are very different ways to model uncertainty. The Sure-thing Principle allows no interaction between states in the evaluation of acts. In effect, this means that the evaluation of an outcome is independent of how probable it is. And the Sure-thing Principle is a necessary condition for representing preferences by an expected utility function.

The idea that certainty has a special status, and more generally that the evaluation of an outcome may depend on how probable it is, is a violation of the Sure-thing Principle, because the evaluation of an outcome in one state depends on what happens in other states. Therefore, if uncertainty is modeled based on these features, then preferences cannot be represented by an expected utility function.

Savage next defines *weak preference among outcomes*: outcome  $a$  is weakly preferred to outcome  $b$  if, and only if, act  $A$ , having  $a$  as the outcome in every state, is weakly preferred to act  $B$ , having  $b$  as the outcome in every state. Moreover, an event is impossible if it does not influence the preference among any acts.

*Axiom 3* says that if  $A$  is an act, which has the outcome  $a$  in every state, and  $B$  is an act, which has the outcome  $b$  in every state, and  $X$  is not an impossible event, then  $A$  is weakly preferred to  $B$  given  $X$  if, and only if,  $a$  is preferred to  $b$ .

This means that knowledge of an event  $X$  cannot change a preference among outcomes. In other words, if an act has the same outcome in every state, the outcome should be certain in the sense that it is independent of which state obtains. The description of an outcome should thus contain all relevant information leaving no uncertainty for the decision maker. If the evaluation of some proposed outcome turns out not to be independent of whichever state obtains, it should be regarded as an act, not an outcome.

## Qualitative and Quantitative Subjective Probability

The remaining axioms are concerned with probabilities (with the exception that axiom 6 also has implications for the evaluation of outcomes). Since I am not concerned with probabilities per se, I shall only present the axioms briefly. First some definitions:

A person is offered a *prize* in case  $X$  obtains, if he is offered an act of which he prefers the outcome given  $X$  obtains to the outcome given  $X$  does not obtain.

Consider the following situation, where  $a$  is preferred to  $a'$  and  $b$  is preferred to  $b'$  (☞ [Table 16.3](#)).

Now a definition of *qualitative subjective probability*:  $X$  is at least as probable as  $Y$  if, and only if,  $A_X$  is weakly preferred to  $A_Y$ . The decision maker prefers  $a$  to  $a'$ , and he therefore prefers the act which is more likely to result in  $a$ , rather than  $a'$ . In other words, he considers  $X$  more probable than  $Y$ .

*Axiom 4* mirrors the Sure-thing Principle for qualitative probability. It says that if  $A_X$  is weakly preferred to  $A_Y$ , then  $B_X$  is weakly preferred to  $B_Y$ . Hence, this axiom states that the evaluation of probabilities of states is independent of the prizes offered. The Sure-thing Principle is an independence requirement for the evaluation of outcomes (outcomes can be evaluated independently from the probabilities of states), and this is an independence requirement for the evaluation of probabilities (probabilities of states can be evaluated independently from the outcomes in states).

*Axiom 5* is the rather innocuous assumption that there is at least one pair of outcomes,  $a$  and  $b$ , such that  $a$  is preferred to  $b$ . If axiom 5 were false, then the relation – is more probable than – would be empty.

☞ [Table 16.3](#)

Illustration of qualitative probability and axiom 4

	$X$	$\sim X$		$Y$	$\sim Y$
$A_X$	$a$	$a'$	$A_Y$	$a$	$a'$
$B_X$	$b$	$b'$	$B_Y$	$b$	$b'$

*Axiom 6* says that if  $A$  is preferred to  $B$ , and  $x$  is any outcome, then there exists a partition of the set of states such that, if  $A$  is modified in any event of the partition to take the value  $x$  in all states there, other outcomes being the same, then modified  $A$  remains preferred to  $B$ ; if  $B$  is modified in any event of the partition to take the value  $x$  in all states there, other outcomes being the same, then  $A$  remains preferred to modified  $B$ .

Consider first the case in which  $A$  is modified.  $A$  is modified in some subset of states by replacing the original outcomes in these states by the outcome  $x$ . The axiom requires that modified  $A$  remains preferred to  $B$ . If  $x$  is a better outcome than the original unmodified outcomes, this is trivially fulfilled. Thus, the axiom only has teeth in the case when  $x$  makes  $A$  worse. In this case, the axiom implies that no matter how bad an outcome  $x$  is, there exists a partition of the set of states into events that are sufficiently fine grained to make the probability of any of them so close to zero that if  $A$  is modified with  $x$  in any of them, modified  $A$  is still preferred to  $B$ . And similarly, no matter how good an outcome  $x$  is, a partition of the set of states into events exists such that if  $B$  is modified with  $x$  in any of them,  $A$  is still preferred to modified  $B$ .

This is a quite strong assumption. Firstly, it ensures that one can always find an event, which corresponds to each probability number, such that the probability measure is unique. Secondly, it is also serves as a continuity condition. In measurement theory, such conditions are often called *Archimedean* conditions, because they correspond to the *Archimedean* property of real numbers: for any number  $x$ , no matter how small, and any number  $y$ , no matter how large, there is an integer  $n$  such that  $nx \geq y$ , i.e., the ratio between any two numbers is always finite. Axiom 6 excludes outcomes, which are “infinitely” good or bad. If there was an infinitely bad outcome  $x$ , modified  $A$  would not remain preferred to  $B$ , regardless of how low the probability of  $x$  occurring; and  $x$  would have to be assigned an infinitely low utility; if there were an infinitely good outcome  $x$ ,  $A$  would not remain preferred to modified  $B$ ; and  $x$  would have to be assigned an infinitely high utility. Thus, axiom 6 ensures that any outcome can be assigned a real-value utility number (i.e., utility fulfills the *Archimedean* property of real numbers).

## The Representation and Uniqueness Theorems

---

From the 6 axioms follows the conclusion:

There exists a probability measure  $p_i(s_i)$  assigning probabilities to states that sum to one, and a real-valued utility function  $u(a_i)$  assigning utilities to outcomes

$$u(A) = p_1 u(a_1) + p_2 u(a_2) + \dots + p_s u(a_s) \text{ such that}$$

$A$  is weakly preferred to  $B$  if, and only if,

$$p_1 u(a_1) + p_2 u(a_2) + \dots + p_s u(a_s) \geq p_1 u(b_1) + p_2 u(b_2) + \dots + p_s u(b_s); \text{ and}$$

the probability measure is unique, and the utility function is unique up to linear transformation.

I shall not present a proof. Savage's own proof is quite complicated. A more accessible proof is probably Fishburn (1970, pp. 191–214).

## The Maxim “Maximize Expected Utility”

---

Many expositions of decision theory are summarized in the maxim that the decision maker should maximize expected utility. Gärdenfors and Sahlin (1988, pp. 1–5) is a telling example. Introducing Savage’s decision theory, they say that the assumption is that the decision maker *evaluates the possible outcomes by assigning a utility measure to them*, and that he possesses probability measures of the likelihood of each state. He is then supposed to *compute the expected utility* of the various alternative acts and choose the one with the maximum expected utility. Gärdenfors and Sahlin claim that Savage’s representation and uniqueness theorems provide the foundation for these assumptions and the principle of maximizing expected utility.

Although popular, it should by now be clear that this is a misunderstanding of the theory. It is true that Savage’s theorems provide the decision maker with a cardinal utility measure for outcomes and a unique probability measure for the likelihood of states. It is also true that a decision maker, who has a cardinal utility measure for outcomes and a unique probability measure as given, and who forms his preferences according to a computation of the expected utility of acts, trivially satisfies Savage’s axioms. Still, Gärdenfors and Sahlin put things upside down; they begin with numerical evaluations of outcomes (called “utilities”) and probabilities and let the preferences over acts be the result of a calculation of the expected utility of the alternative acts. But theory starts out from the assumption that the decision maker has *preferences over pairs of acts*. It then proves that if these preferences over acts conform to the axioms, they can be represented by a function with an expected utility form. Preferences over acts are the primitives of the theory, from which the numerical representation of outcomes is deduced.

Utility is a theoretical construct that summarizes the information already given by the preferences. To say that an individual maximizes his expected utility is just to repeat that he chooses according to his preferences. As it is often carefully expressed, if the decision maker’s preferences satisfy the axioms, his choices can be described *as if* he maximized (the theoretical construct) utility (e.g., Luce and Raiffa (1957, p. 31)). But utility is *not* defined as a substantial objective; it is not assumed to play any role in the decision maker’s deliberations as something he aims to maximize. The decision maker does not prefer one act to another *because* it has a higher utility, he simply prefers it and therefore it has a higher utility.

The difference is not only about the role of utility in deliberation, it is also about the decision maker’s attitude to risk: If the decision maker assigns “utility” to an outcome according to the *value* for him of the outcome occurring, maximizing expected utility implies *risk neutrality about this value*. Preferences based on this calculation therefore ignore the question of risk. If utility, on the other hand, is derived from basic preferences over acts, these preferences are defined to take risk into account, and utility therefore embodies the considerations of risk. The utility of an outcome cannot be identified with the value for the decision maker of the outcome.

Hence, the central misunderstandings are, firstly, not to recognize that *preferences over acts are assumed by the theory* and not derived from it; moreover, not to recognize that *utility is defined to represent these assumed preferences over acts* and, even though the theory assigns utilities to outcomes, these utilities are derived from the basic preferences over acts and thus reflect the decision maker’s attitude to risk. I stress these points because they form a recurrent theme in this chapter.

## The Problem of Action Guidance

---

However, if decision theory simply assumes that the decision maker has a complete preference ordering for acts, then he already knows what to do. It seems a reasonable expectation that decision theory should guide us in dealing with uncertainty. Now we are told that it simply assumes that we already know what to do. What is the use of a theory of this kind? The first step in answering this question is to make the message of decision theory clear. What does it recommend?

Well, taken literally, it does not recommend anything. The theory consists of a representation theorem and a uniqueness theorem, which are two statements of the “if . . . , then . . . ” form. It does not contain any prescriptions. However, the standard interpretation of theory assumes it to have normative implications.

Decision theory (in Savage's version) is born with the following interpretation: the empirical relation is the decision maker's weak preference relation, and it is defined on a set of objects which is the product set of the alternative acts and the set of the possible states; more precisely, the preferences are defined on the rows of this product set, i.e., on acts understood as uncertain prospects, consisting of the outcome in each possible state ( $a_1, a_2, \dots, a_s$ ). The numerical representation assigns nonnegative numbers that sum to 1 to the states (called “probabilities”) and real numbers (called “utilities”) to outcomes in such a way that one act is preferred to another if, and only if, it has a higher expected utility.

On this standard interpretation, the axioms are understood as requirements of rationality for choice behavior. These requirements are only formal. They are not about the content of preferences, but about consistency requirements between them. The claim that it is rational to conform to the axioms has been backed up by additional theorems aiming to show that a decision maker, who violates the axioms, risks ending up with less preferred outcomes than he could have achieved. He may thus fare worse by his own standards than he could have, and to end up choosing a less preferred outcome when a more preferred outcome could have been obtained is clearly not rational.

One such theorem says that an individual who violates transitivity may be exploited by a so-called money pump. A violation of transitivity implies cyclical preferences (e.g.,  $A$  is preferred to  $B$ , which is preferred to  $C$ , which is preferred to  $A$ , and so on). Assume that you are willing to pay a small amount for obtaining a more preferred outcome. If you have cyclical preferences, then, at some stage, you will end up with the outcome with which you started, but this time with less money (Davidson et al. 1955). Another theorem says that an individual who violates the laws of probability (which are implied by the axioms) may be exploited by a so-called Dutch Book. A Dutch Book is a combination of bets that is certain to lead to a loss. For instance, if your probabilities, in a given case, do not sum to 1, this fact can be exploited by setting up bets you will be likely to accept because of your probabilities, but which taken together will lead to a loss for sure (Kemeny 1955; Lehman 1955; Shimony 1955).

The strength of this kind of justification for the axioms (often called *pragmatic arguments*) is contested (Schick 1986; Rabinowicz 1995). Also, axioms like completeness or the Sure-thing Principle are in need of another kind of justification. The Sure-thing Principle was discussed briefly above. And why should completeness be a requirement of rationality? Or axiom 6 for that sake? However, again I shall not pursue these questions any further, because the interest here is to take the axioms for granted and to ask what we can learn from them with regard to decision making under uncertainty.

Hence, if we take the message of decision theory to be that preferences should conform to the axioms, we shall have to take a closer look at the axioms to get a grip on what is recommended. The axioms are conditions for the structure of preferences. They do not tell us what preferences to have, but once we have some preferences, they constrain the additional preferences we could rationally have together with them.

The most important axiom for decision theory's handling of uncertainty is no doubt the Sure-thing Principle. It tells us that, when comparing two acts, we should compare them state-by-state and then weigh up the differences between the possible outcomes in such a way that they can be weighted with any kind of probability. Utility assignments embody this weighing up. Consider  [Table 16.4](#) below (taken from Broome (1991b, p. 146)), the numbers represent income per week for some individual.

It follows from the theory that  $A$  is preferred to  $B$  if, and only if,  $u(A) > u(B)$ , i.e., if  $\frac{1}{2}u(100) + \frac{1}{2}u(200) > \frac{1}{2}u(20) + \frac{1}{2}u(320)$ , which again is equivalent to  $u(100) - u(20) > u(320) - u(200)$ . These utility differences tell us how much the difference between these outcomes counts. The risk of getting €20 instead of €100 counts for more than the chance of getting €320 instead of €200. Had it been the other way round,  $B$  would have been preferred to  $A$ .

Weighing up these differences *determines the attitude to risk*, and the *considerations concerning risk are embodied in the utility assigned to each possible outcome*. Note that the utility assigned to an outcome is independent of probabilities. The weighting embodied in the utility assigned to outcomes is supposed to work for any probability distribution over the states. In the example, the states are equally probable, but the same weighting should cover all possible unequal distributions as well.

Remember that Bernoulli coupled the claim about risk aversion about money with the claim that money has diminishing marginal value. However, as pointed out by Hansson (1988), the diminishing marginal value of money and risk aversion about money gambles are conceptually two different things. So, if the decision maker is risk averse about money gambles, the utility of an outcome is not just the value for him of the outcome, but rather an increasing strictly concave transform of the value for him of money. I return to this point below.

To summarize, decision theory tells us to have preferences that conform to the axioms. From the structure these axioms impose on preferences, we can derive some recommendations. Most importantly, the Sure-thing Principle tells us to form preferences by comparing differences between outcomes state by state. Likewise, recommendations for forming subjective probabilities can be derived, but I shall leave them aside here.

 **Table 16.4**  
Weighing outcomes in the face of risk

Prospect A		Prospect B	
State heads $p = \frac{1}{2}$	State tails $p = \frac{1}{2}$	State heads $p = \frac{1}{2}$	State tails $p = \frac{1}{2}$
€100	€200	€20	€320

## Decision Theory as a Theory of Good

### External Reasons

According to the standard interpretation, the rationality requirements are purely formal. The content of the decision maker's beliefs about probabilities, and his preferences over acts, seems to be exempted from the requirements of rationality. To put it another way, the theory does not tell us how we should form our beliefs and preferences; it only sets up some restrictions for this formation process. However, even though it is not part of the theory itself, the standard interpretation will typically refer to *Bayes' Theorem*. This theorem shows, roughly, how we can revise initial probabilities in the light of new information (again, I cannot go into details here, but see Peterson (2009, pp. 125–130) for an introduction and Mellor (2005) for a further discussion). This indicates that, even on the standard interpretation of decision theory, there is some interest in justifying beliefs by reasons that are external to the theory.

Conventional Bayesian decision theorists will, of course, insist that an agent is free to believe and prefer whatever he wishes, as long as he fulfills the formal rationality requirements. However, others have advocated some form of combination of the formal apparatus of decision theory with substantial external theories about what to believe and prefer. For one thing, it is possible to combine decision theory with a more objective approach to probabilities (Broome 1991b). For the rest of this chapter, I shall assume such an approach without being specific about it. I shall simply assume that there is one “best” set of probabilities, which applies to all individuals.

My main concern will be with another concept: the idea that some external substantial theory guides us when determining what is good for an individual, or good overall. It is a reasonable assumption that a theory about what is good for an individual in the face of uncertainty, or a theory about what is good overall in the face of uncertainty, should satisfy the rationality requirements of decision theory, and this is the idea I shall investigate further. The substantial content should come from a theory about what is good for an individual, or from a theory about what is good overall.

In this vein, John Broome (1991b) has suggested an interpretation of decision theory as a theory about good. He assumes the existence of a weak betterness relation for each individual (— is at least as good for individual  $i$  as —) and a weak general betterness relation (— is at least as good as —) defined over acts (i.e., uncertain prospects). It is understood that theories about self interest and ethics determine how the individual relations and the general relation order the acts. On this interpretation, decision theory implies that, if a weak betterness relation (be it personal or general, I shall often discuss these cases together) satisfies the axioms, then it can be represented by an expected utility function.

On the face of it, decision theory is here connected to axiology, i.e., the study of values – more precisely the study of what is good for individual and what is generally good. However, for Broome, the general betterness relation is directly connected to ethics, i.e., the study of what is right to do. Broome defines *teleological ethics* by the claim that the right act is determined by an ordering according to betterness. Standard teleology has a *maximizing structure*: it says that if one act is higher in the ordering then that act is right, and all other acts are wrong.

To see matters this way is rather rare among philosophers. It is more common among economists who do similar things in welfare economics under uncertainty. However,

economists, unlike philosophers, are not used to the concept of goodness. They generally prefer to talk about individual preferences and social preferences. John Broome was originally an economist, but he became a philosopher; and he has thereby been able to bridge the two traditions. A specific point for Broome has been to avoid certain problems with subjective probabilities in welfare economics or social choices by a more objective approach (see more below).

Among philosophers, it is much more common to refer to the maxim “maximize expected goodness.” Derek Parfit (1984) says: “What we ought subjectively to do is the act whose outcome has the greatest *expected* goodness” (p. 25). Michael Zimmerman (2008) advises people to do the act which is *prospectively better*, and “prospectively best” is to be understood in terms of expectable value for the agent” (p. 126).

The widespread opinion among philosophers is that decision theory *implies* the maxim of maximizing expected goodness. Some moral theory provides us with an ordering of outcomes according to their moral goodness. And decision theory supposedly tells us to choose the act, which leads to the highest expectation of goodness. Frank Jackson (1991) says (pp. 463–464):

- ▶ Generalizing, the proposal is to recover what an agent ought to do at a time according to consequentialism from consequentialism’s value function – an assignment of value that goes by total consequent happiness, average consequent preference satisfaction, or whatever it may be in some particular version of consequentialism – together with the agent’s subjective probability function at the time in question in the way familiar in decision theory, with the difference that the agent’s preference function that figures in decision theory is replaced by the value function of consequentialism. That is to say, the rule of action is to maximize  $\sum_i Pr(O_i/A_j) \times V(O_i)$ , where  $Pr$  is the agent’s probability function at the time,  $V$  is consequentialism’s value function,  $O_i$  are the possible outcomes, and  $A_j$  are the possible actions.

For Jackson, “the agent’s preference function” is the agent’s ranking of states of affairs, and it is supposed to be replaced by consequentialism’s ranking. But, as we have seen, decision theory does not advocate the maxim of maximizing expected goodness. However, apart from the mistaken reference to decision theory, philosophers might have other reasons for seeing things this way.

First, Broome’s interpretation of decision theory presented here considers the goodness (or rather betterness) of uncertain acts basic. But when we determine whether one uncertain act is better than another, we do it relative to the available probabilities. This sort of probability-relative betterness appears as an interim sort of goodness. What ultimately matters is what actually happens, once the uncertainty is resolved. Hence, it could be argued that goodness is truly a property of final outcomes and not a notion that can be applied on the basis of probabilities. Most ethical theories have in fact discussed goodness as a property of outcomes with no uncertainty involved, i.e., they assume the ideal situation of choice under certainty. Secondly, the maxim “maximize expected goodness” might be considered more action guiding than simply assuming a complete ordering of acts.

However, both of these ideas have serious problems of their own when it comes to dealing with uncertainty. Ironically, these problems arise exactly because the idea of goodness as a property of final outcomes and the connected maxim of maximizing expected goodness overlook the central insight of decision theory.

## Probability-Relative Goodness

---

Consider first the former claim. Everybody seems to agree that if there is uncertainty at the time when the decision is to be made (and decision theorists would insist that there is *always* uncertainty), the decision maker has to base his decision on the available probabilities. On the other hand, everybody also agrees that what ultimately matters is the final outcome.

Often, it is put in the way that *subjectively* (cf. the Parfit quote above) we ought to do the act whose outcome has the greatest expected goodness. But later on, this act can turn out to be *objectively* wrong. The challenge is: How can we say that one act is better than another on the basis of probabilities when later on, once the uncertainty has been resolved, the act may not turn out to be better after all? What is really the right thing?

Philosophers have struggled with this question. G. E. Moore was perhaps the most firm advocate of the view that goodness is a property of final outcomes. He drew the conclusion that, in the face of uncertainty, you can never know what is the right thing to do (Moore 1903, p. 199). But this would imply that ethics has nothing to say about how to act under conditions of uncertainty.

Using an example set up by Frank Jackson (1991, pp. 462–463), Michael Zimmerman has argued that there are cases in which Moore's view does not imply a lack of knowledge about what is right under conditions of uncertainty, but actually leads to the wrong conclusion:

- ▶ Jill is a physician who has to decide on the correct treatment for her patient, John, who has a minor but not trivial skin complaint. She has three drugs to choose from: drug A, drug B and drug C. Careful consideration of the literature has led her to the following opinions. Drug A is very likely to relieve the condition but will not completely cure it. One of the drugs B and C will completely cure the skin condition; the other though will kill the patient, and there is no way she can tell which of the two is the perfect cure and which is the killer drug. What should Jill do?

(In Zimmerman's account, A is called *B*, and B is called *A*). A decision theoretic analysis of the case looks like  [Table 16.5](#).

In this case, Moore knows that A results in an outcome which is *not* the best. He is therefore forced to say that this act is wrong. But that seems clearly to be the wrong conclusion (Zimmerman 2008, p. 18). B and C are too risky, and Moore's view does not seem to be able to take risk into account.

 **Table 16.5**  
**The Jackson case**

Acts \ States	B completely cures the condition C kills the patient $p = \frac{1}{2}$	C completely cures the condition B kills the patient $p = \frac{1}{2}$
Drug A	Partial cure	Partial cure
Drug B	Complete cure	Patient dies
Drug C	Patient dies	Complete cure

Clearly, the subjective view that whatever the decision maker believes is right (i.e., whatever subjective probabilities he has) is actually right will not do either. It is of course still the case that what ultimately matters is what actually happens. The way to deal with this fact is to make a distinction between probabilities of different status (Broome 1991b, pp. 126–131). Objective probabilities (if they exist) have a higher status than subjective probabilities. Probabilities based on more evidence have higher status than probabilities based on less. At the limit, what would actually happen has the highest status. I cannot present a detailed account of the status of probabilities here, but I hope the idea at least seems intuitively plausible.

You ought to base your judgment on the best available probabilities. Looking at the Jackson Case, we know that *A* does not lead to the best possible outcome, but either *B* or *C* does. Given the best available probabilities, it would be too risky to choose one of these acts. It is better to choose *A*. The certain knowledge we have in this case does not influence the evaluation of the acts (given the available probabilities).

There is a thorough discussion of these questions in Zimmerman (2008). He advocates what he calls the Prospective View, according to which, very roughly, we ought to evaluate the acts on the basis of the best available probabilities. The details of Zimmerman's theory about uncertainty are rather complex, and in some ways they are at odds with decision theory according to Savage's version; however, I cannot go into these details here. One thing worth mentioning, though, is that Zimmerman still believes goodness to be a property of final outcomes, and he understands the Prospective View in terms of the maxim "maximize expected goodness." This prevents him from fully acknowledging the insights of decision theory, with regard to dealing with uncertainty.

## Problems for the Maxim of Maximizing Expected Goodness

---

The maxim of "maximizing expected goodness" faces two severe problems. The first problem is a problem of measurement. We know that the operation of forming the mathematical expectation of some quantity requires that the quantity is cardinally measurable, i.e., it must be determined uniquely, at least up to linear transformation. But it is doubtful whether the goodness of outcomes can be measured with so much numerical structure. The standard assumption in ethical theory is that there is an ordering of outcomes in terms of their goodness. And the lesson from economics was exactly that a simple ordering of outcomes can only be represented by an ordinal scale.

It might be thought that decision theory supplies us with a cardinal measure of the goodness of outcomes. But this is not true. Decision theory provides us with a cardinal measure of the *utility* of outcomes. To identify utility with goodness is a mistake. "Utility" is defined as a function, which represents the betterness relation. In doing so, it assigns numbers (utilities) to outcomes in possible states in such a way that one act is at least as good as another if, and only if, it has at least a high expected utility.

The utility of an outcome reflects its weight in the comparison of uncertain *acts* in terms of betterness. There is no direct way to separate out the *goodness of the outcome* in itself, so to speak. In fact, it is difficult to specify what the goodness of an outcome in itself means. The measurement of goodness is determined by some context. The context of decision theory is the comparison of acts in terms of betterness. Perhaps goodness is measured by the comparison of pure outcomes in a context with no uncertainty? But comparing outcomes in terms of betterness in a context of certainty can only lead to ordinal measurement of goodness.

At any rate, the utility of outcomes is not an entity which it makes sense to use for calculations of expected utility, because it is already defined as that which the decision maker maximizes the expectation of. There is a complete ordering of acts, and utility represents this ordering. To calculate expected utility is just to repeat what we already know.

The challenge for the maxim of maximizing expected goodness is therefore to find a cardinal measure of the goodness of outcomes, which is independent of decision theory, but which at the same time orders outcomes in the same way. The only way to obtain this is through modeling the domain of outcomes with a richer structure that would allow numerical representation by a cardinal scale. This might not be impossible. However, the best known suggestions of this sort depend on strong premises and are considered controversial (there is an example in Peterson (2004); see also the discussion in Broome (1991b); Jensen (1995); Ellingsen (1994)).

But suppose for a moment that the goodness of outcomes could be measured cardinally (independent of decision theory). The second problem is then that the maxim of maximizing expected goodness implies *risk neutrality about good*. Only the expectation of goodness counts, it has no value to avoid risk. This appears rather implausible. Consider a choice between 100 units of goodness for sure, and an option that will lead, with equal probability, to either no units, or 200. These options have the same expectation of goodness, but it seems clearly better to play safe and avoid the risk of getting nothing at all.

At any rate, the maxim of maximizing expected goodness assumes risk neutrality about goodness by definition, so to say, as if it was a direct implication of decision theory. Remember that Bernoulli advocated risk aversion about money, but *Bernoulli's Hypothesis* was exactly to maximize the expectation of goodness, and this implies risk neutrality about good. This may be the reason why many people believe that decision theory implies risk neutrality about good. Also on this point, a fundamental insight of decision theory is overlooked, because decision theory allows us to keep the question about the attitude toward uncertainty about goodness open, simply by recognizing that utility is *not* identical to goodness with certainty.

On the face of it, risk aversion about goodness appears very reasonable. Indeed, if there is reason to be risk averse concerning money, as many people are, there is so much more reason to be risk averse about goodness. It certainly requires an argument to justify risk neutrality about goodness.

It is clear that decision theory and the goodness measure should agree on the *ordering* of outcomes. If utility, as determined by decision theory, happened to be a linear transform of goodness, decision theory would imply risk neutrality about goodness. However, since some form of risk aversion is more plausible, utility is likely to be an increasing nonlinear transform of goodness, more precisely an increasing concave transform of goodness (see ➤ Fig. 16.2).

---

## Bernoulli's Hypothesis

Broome (1991b) puts forward the proposal that we consider utility as a measure of goodness (pp. 142–148, 213–222; see also Jensen (1995)). This would answer both problems: There is now a cardinal measure of the goodness of outcomes, and since utility is defined as that which the decision maker maximizes, risk neutrality about good follows automatically. Broome claims that this would amount to a modern version of *Bernoulli's Hypothesis*: an individual ought to maximize the expectation of his good.

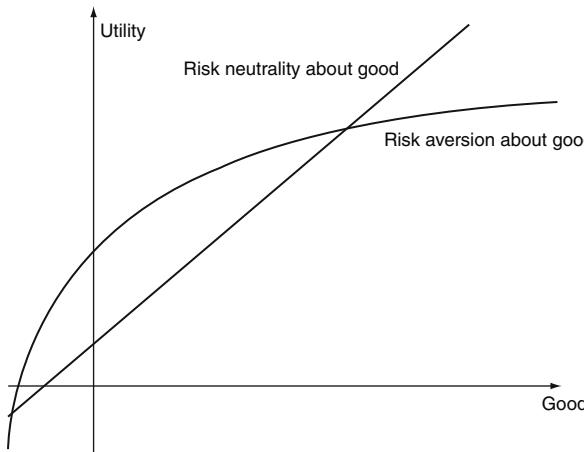


Fig. 16.2

### Utility as a transform of goodness

Broome's suggestion is that a distinction between good and utility (i.e., how much good counts in decision making) appears empty. "Utility" is simply the most plausible candidate to provide the notion of a cardinal measure of the goodness of outcomes with meaning.

But this reasoning seems dubious. The standard meaning of the goodness of an outcome is its value as a certain outcome for the individual. The utility of an outcome, on the other hand, is the strength of the reason in favor the occurrence of this outcome has in decision making under uncertainty, regardless of the probability with which it occurs. Utility summarizes all relevant information about the attitude to risk, whereas the goodness of an outcome contains no reference to risk. These are clearly different concepts.

Hence, Broome's suggestion does not capture the standard meaning of goodness. So when he says that his modern version of *Bernoulli's Hypothesis* implies "risk neutrality about good," this is not risk neutrality about good in the ordinary sense. The question of whether an individual should be risk neutral about good in the ordinary sense remains controversial. But Broome is of course right that the context in which goodness can be measured cardinally remains unclear.

To summarize, the maxim of maximizing expected goodness leaves no room for risk aversion about goodness. Only a betterness relation defined on uncertain acts leaves room for a choice between risk neutrality and risk aversion about goodness. Or rather, a betterness relation defined on acts leads to a utility function, which embodies considerations of goodness and risk simultaneously, where there is no direct way to separate these considerations. This is why I call it a core insight of decision theory to consider the ordering of acts as basic rather than the ordering of outcomes.

## General Good and Individual Good

I shall now move on to the realm of ethics and consider the weak general betterness relation. More specifically, I shall be concerned with how the weak general betterness relation relates to the individual weak betterness relation. The weak general betterness relation implies some

attitude to risk for general betterness. It may seem a natural suggestion that attitude to risk implied by general betterness' should be a function of the individual betterness relations and the attitude to risk implied by them. In other words, I shall examine the idea that general betterness should reflect how uncertain acts look from the personal perspective.

## Equalizing Risk

First, I shall examine the idea that there should be a concern for equality in the prospects people face. Peter Diamond (1967) has demonstrated that the Sure-thing Principle has consequences for general betterness, which once again makes it a controversial axiom, because it is in conflict with this concern. Consider Diamond's example in [Table 16.6](#). There is a kidney available for transplantation and two people  $P$  and  $Q$  who need a kidney in order to survive. The choice is between  $A$ , tossing a coin to decide who is to get the kidney, or  $B$ , giving it directly to  $P$ .

Note that this is a boiled-down version of the more general question about risk distribution: Knowing that one individual is going to die, is it better that two individuals have an equal chance of dying, rather than one individual dying for sure? Or if 100 people will die, is it better to have 10 million people exposed to a 1 in 10,000 chance of dying, than to have 10,000 exposed to a 1 in 100 chance?

The Sure-thing Principle for general betterness implies that we should judge what happens in one state, given fixed outcomes in the other, independently of what these other outcomes are. Suppose we judge it better in state Heads that  $P$  lives and  $Q$  dies, given that in state Tails,  $P$  dies and  $Q$  lives. According to the Sure-thing Principle we should retain this judgment of what happens in state Heads, also given that in state Tails,  $P$  was to live and  $Q$  was to die. If we do, then  $A$  and  $B$  are equally good. But if we believe that people should have equally good prospects, then we should deny this implication. Hence, equalizing risks violates the Sure-thing Principle for general betterness.

Clearly, at least as long as there is nothing to discriminate between these individuals, it appears very unfair to simply pick out one of them to die. The trouble is that two different conclusions can be drawn from this intuition. One is to opt for the general view that there is value in equalizing the prospects people face. This option involves giving up the Sure-thing Principle for general betterness. In this option, the unfairness is not located in any particular outcome. The unfairness is only visible when we compare what happens in one state with what happens in another. Allais wanted to give special weight to certainty about some preferred outcome. Diamond wants to give special weight to avoid the certainty of a bad outcome, and more generally, to equalize the prospects people face.

**Table 16.6**  
Diamond's case

	Act A		Act B	
	State heads $p = \frac{1}{2}$	State tails $p = \frac{1}{2}$	State heads $p = \frac{1}{2}$	State tails $p = \frac{1}{2}$
Individual $P$	Lives	Dies	Lives	Lives
Individual $Q$	Dies	Lives	Dies	Dies

The other option, however, is to say that if there is value in avoiding unfairness, this value should be located in some outcome. Hence, it could be said that the unfairness is indeed done to individual Q in act *B*; and therefore it should be located as part of the outcome in both states here: “dies and treated unfairly.” This is the line John Broome (1991b, pp. 110–115) takes. It avoids violating the Sure-thing Principle, because there is now a difference between the outcomes for Q in the two prospects.

More generally, Broome rejects the idea that there is any merit in equalizing the prospects people face. The reason is simply that if equality is valuable, it is equality in the final outcome that matters, not equality in the prospects people face. This is compatible with the observation that if there is not enough of some indivisible good to treat all people equally, such as the kidney in Diamond’s example above, we owe people equal chances.

## The Principle of Personal Good

---

Assume now that general betterness and individual betterness satisfy the axioms of decision theory. Next, I shall discuss the general idea that the attitude to risk in general betterness should reflect what is good for individuals in the face of uncertainty.

The *Pareto Principle* connects general good (which economists call “social preferences”) with individual preferences: Two acts are socially indifferent if everyone is indifferent between them; and if everyone weakly prefers one act to another, and at least one strictly prefers it, then it is socially preferred. Applied to uncertain acts and allowing for subjective probabilities, the Pareto Principle conflicts with Savage’s axiom 4 for “social preferences.” I cannot go into details here (they are explained in Broome (1991b, pp. 152–163), but it should be intuitively understandable that if people disagree about probabilities, it will be impossible to determine “social probabilities” for “social preferences.”

In the face of this conflict, some economists stick to the Pareto Principle and give up decision theory for “social preferences”; this is called the *ex ante* approach to the aggregation of preferences. But since the task here is to examine the consequences of accepting decision theory, I shall look in another direction. The *ex post* approach to the aggregation of preferences sticks to decision theory and instead abandons the Pareto Principle for acts (but retains it for outcomes). This approach builds on a distinction between preferences over acts, which involves beliefs about probabilities, and preferences over outcomes which presumably does not. The *ex post* approach wants to skip the former and use “social probabilities” instead, but keep the latter as the basis for aggregation.

But this is a misleading distinction. In decision theory, preferences over acts are basic, and they depend on beliefs about probabilities. Preferences over outcomes are derived from these preferences over acts. Hence, there is no way to find pure preferences over outcomes that are independent of beliefs.

Instead, Broome (1991b) has suggested what he calls the *Principle of Personal Good* for uncertain acts. This principle assumes that the same probabilities apply for all individuals, as well as for general good. This is in line with the *ex post* approach. But it is more objective in the sense that it frees the evaluation of an individual’s good from his preferences (which are based on his subjective beliefs).

- ▶ *The Principle of Personal Good:* Two alternatives are equally good if they are equally good for each person. And if one alternative is at least as good as another for everyone, and definitely better for someone, it is better.

The idea underlying the Principle of Personal Good is that general betterness should be responsive to the betterness of an individual. Thus, it is more in spirit with the original Pareto Principle insofar as it insists on the perspective on uncertainty from a personal point of view as being fundamental for the aggregation of betterness across individuals. But the Principle of Personal Good is only sparsely discussed in the literature.

The Principle of Personal Good is defined for uncertain acts and thus implies a notion of probability-relative good. Hence, it implies that general betterness should be responsive to the betterness of individuals, given the available probabilities. But this may lead to a conflict with the concern that general betterness should be responsive to the betterness of individuals in final outcomes.

Consider the following example (☞ *Table 16.7*) set up by Broome (1991b, pp. 170). The states are equally probable. The figures in the table stand for the income of two individuals. Assume that the good for these two individuals is determined by their income only, and assume that it is good for both individuals to avoid risk to their income.

For both individuals, the two acts have the same expectation of income, but *B* is risky, whereas *A* is not. Given the assumptions, *A* is better for both individuals, and the Principle of Personal Good therefore implies that *A* is generally better.

These judgments are relative to the available probabilities. However, we know for sure that whatever state occurs, one or the other individual will end up worse under *A* than he/she would have been under *B*. Relative to what actually happens, *A* is *not* better than *B* for both people. The question is whether we can trust the Principle of Personal Good under these circumstances and declare *A* to be generally better than *B*, given the probabilities.

Broome has raised this question. Initially, he thought the answer was “no, we cannot trust the Principle of Personal Good” (Broome 1978). However, later Broome (1991b) came to believe that there was a defense of the probability-relative Principle of Personal Good. He says (p. 172):

- *A* is better for *P* than *B* because *A* is less risky, and I assumed it is good for *P* to avoid risk. In expected utility theory, the goodness of avoiding risk appears in the form of weights attached to gains and losses. Here it gives the difference between 2 and 1 more weight than the difference between 3 and 2. [...]

[...] the fact that it is good for *P* to avoid risk is something that must contribute to determining the general goodness of alternatives. Since this goodness appears in the weighting of reasons, the weighting must be preserved in determining which alternative is generally better. This is why *A* is generally better than *B*. The benefit of risk-avoidance to the two people must appear in general good.

■ **Table 16.7**  
A problem for the Principle of Personal Good?

	Act A		Act B	
	State heads $p = \frac{1}{2}$	State tails $p = \frac{1}{2}$	State heads $p = \frac{1}{2}$	State tails $p = \frac{1}{2}$
Individual <i>P</i>	2	2	1	3
Individual <i>Q</i>	4	4	5	3

The trouble is that this defense does not work as it stands. From the point of view of general betterness, in order to evaluate  $A$  and  $B$ , we have to compare them state by state. In state Heads,  $P$  gets 2 in  $A$  and 1 in  $B$ , whereas  $Q$  gets 4 in  $A$  and 5 in  $B$ , i.e.,  $A$  is better for  $P$  and  $B$  is better for  $Q$ . In state Tails,  $P$  gets 2 in  $A$  and 3 in  $B$ , and  $Q$  gets 4 in  $A$  and 3 in  $B$ , i.e.,  $A$  is better for  $Q$  and  $B$  is better for  $P$ . There is a conflict of interest between  $P$  and  $Q$  in both states.

How should general betterness take risk into account? Risk neutrality about good seems as implausible for general betterness, as for individual betterness. The utility of an outcome should reflect the attitude to uncertainty in general betterness. Broome also assumes that the Principle of Personal Good applies for outcomes. It seems to follow that general utility for an outcome should be an increasing function of what is good for the individual in the face of uncertainty (as measured by the individual's utility function). And Broome actually suggests that we should use the weights derived from the individual betterness relations, i.e., utilities, when taking the benefit of risk avoidance for the individuals into account in general betterness.

Given the assumption that it is good for the individuals to avoid risk to income (and goodness), their utility functions will be increasing and strictly concave. However, an increasing, strictly concave transformation of the incomes in [Table 16.7](#) does not resolve the conflict of interest between  $P$  and  $Q$  in both states. As long as we do not know how to compare betterness across individuals, we cannot resolve the conflict and say whether  $A$  is generally better than  $B$ , or the other way round. Hence, the fact that  $A$  is better for both individuals is *not* a sufficient reason to say that  $A$  is generally better.

The Principle of Personal Good applied to uncertain acts seems a natural way to deal with risk to general betterness. However, risk to individuals can be taken into account by letting the betterness of outcomes be an increasing function of individual utilities in these outcomes. Also, looking at risk from the point of view of individuals conceals that there is a conflict of interest in both states. This is not to say that the Principle of Personal Good for uncertain acts cannot be justified, but the justification requires stronger assumptions than the mere point that considerations of risk to individuals should appear in general good. I shall point out three important implications of this discussion.

## Valuing Life

The first is about valuing life. Before presenting the case above, Broome (1978), reprinted in Broome (1991, pp. 177–182), in another context, had cast a widespread procedure for valuing the life of individuals in doubt. Suppose that some public project is up for consideration. Apart from bringing some benefits, the project will increase the probability of dying for everyone. Further, assume that in the end one individual will definitely die. The procedure is now to compensate people for the increased risk of death in such a way that everyone, at the time when the project is considered under conditions of uncertainty, accepts that they are better off with the project and compensation, than without. See [Table 16.8](#).

If acceptance is granted as evidence for people actually being better off (under the uncertain conditions), then the Principle of Personal Good applies, and the project judged in advance is better than the status quo. However, later on, when the uncertainty is resolved and one specific

individual is dying, he will not be better off, because in the face of certain death, the compensation will appear hopelessly inadequate. Broome says (1978, p. 94):

- ▶ If the justification for accepting a project [...] is that compensation can be arranged so that nobody is harmed, then the justification cannot possibly apply when, after the project has been carried out and the utmost has been done by way of compensation, somebody palpably has been harmed, namely the person who has died.

Later, after presenting his defense of the Principle of Personal Good quoted above, Broome (1991b) withdrew this remark (p. 171, note 4). But if the defense does not work, and the Principle of Personal Good requires stronger arguments, the procedure cannot be justified by reference to this principle. From the point of view of general betterness, we shall have to compare the options state-by-state. Assuming that the case is completely symmetrical, this comparison is the same for all states: The procedure is only justified if the (individual utilities of) benefits and compensation to the  $n-1$  people are sufficient to accept the (individual (dis) utility of) loss of one life, and this is *not* ensured by the fact that every individual is willing to run the risk in advance. The judgment from general betterness involves comparisons of betterness *across* individuals.

## The Interpersonal Addition Theorem

However, the next point is that it is possible to state precisely the condition under which the Principle of Personal Good does not conflict with general betterness. This can be inferred from the implications of accepting the Principle of Personal Good. Consider the *Interpersonal Addition Theorem* (Broome 1991b):

There are  $s$  states and  $h$  individuals. Assume individual weak betterness relations and a weak general betterness relation. The theorem says:

- (P1) Each of the  $h$  individual weak betterness relations satisfies the axioms of decision.
- (P2) The general betterness relation satisfies the axioms of decision theory.

Table 16.8

Evaluation of a project causing the statistical death of one out  $n$  individuals

	Project with compensation				No project			
	$p = 1/n$	$p = 1/n$	...	$p = 1/n$	$p = 1/n$	$p = 1/n$	...	$p = 1/n$
Individual 1	Dies	Benefits and compensation	...	Benefits and compensation	Status quo	Status quo	...	Status quo
Individual 2	Benefits and compensation	Dies	...	Benefits and compensation	Status quo	Status quo	...	Status quo
...	...	...	...	...	...	...	...	...
Individual n	Benefits and compensation	Benefits and compensation	...	Dies	Status quo	Status quo	...	Status quo

## (P3) The Principle of Personal Good:

- (a) Two alternatives are equally good if they are equally good for each individual.
- (b) If one alternative is at least as good as another for everyone and definitely better for someone, it is better.

From these premises follows the conclusion:

- (C) The general weak betterness relation can be represented by an expectational utility function that is the sum of expectational utility functions representing the weak betterness relations of individuals.

Two remarks: The first is that this is Broome's reinterpretation of a famous theorem by John Harsanyi (1955), reprinted in Harsanyi (1976, pp. 6–23). Harsanyi's theorem was stated in terms of individual and social preferences. As I said above, Broome uses betterness relations because (1) the same probabilities apply for all, and (2) he does not want to commit himself to the *Preference Satisfaction Theory* about an individual's good. The second remark is that several authors have pointed out that the premises only allow the conclusion that the general utility function can be written as a *weighted sum* of individual utilities. In order to reach the conclusion (C), i.e., that the weights are all one, an explicit premise about interpersonal comparisons is also needed (e.g., Vallentyne (1993)).

Now we turn to the implication of accepting the Principle of Personal Good along with the axioms of decision theory for the betterness relations. Each alternative will be an uncertain prospect:

$$\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_s),$$

specifying for each of  $s$  states the outcome for each of  $h$  individuals. For each individual  $i$  there is an expectational utility function. We can write the utility values in each state of nature for all individuals in a grid like  [Table 16.9](#).

The Sure-thing Principle for general betterness implies that we compare one alternative with another, state by state. For each outcome, we aggregate utilities across individuals, and then we aggregate across states to reach the final judgment. In the discussion above, it was assumed that in aggregating across individuals in an outcome, the Principle of Personal Good applies to outcomes. Hence, general utility in a state is an increasing function of individual utilities. It was also assumed that general betterness deals with risk by taking individual utilities (reflecting the good of risk avoidance for individuals). If there is no risk aversion over and above the risk aversion for individuals, the general utility of an alternative will be its expected utility, given by the sum of general utility for each state, weighted by the probability of the state; and the general utility of a state is an increasing function of individual utilities for that state.

 **Table 16.9**

Individual utilities for the outcomes in all states

	State 1	State 2	...	State $s$
Individual 1	$u_1(\mathbf{x}_1)$	$u_1(\mathbf{x}_2)$	...	$u_1(\mathbf{x}_s)$
Individual 2	$u_2(\mathbf{x}_1)$	$u_2(\mathbf{x}_2)$	...	$u_2(\mathbf{x}_s)$
...	...	...	...	...
Individual $h$	$u_h(\mathbf{x}_1)$	$u_h(\mathbf{x}_2)$	...	$u_h(\mathbf{x}_s)$

Once the Principle of Personal Good is accepted, we learn from the Interpersonal Addition Theorem that it is possible to reach the general utility of an alternative in two ways with the same result. On the one hand we can add up the individual utilities for each state of nature, and then form the general expected utility. But we can also, for each individual, form the expected utility, and then add up across individuals. Hence, an implication of the theorem is that the order of aggregation does not matter.

Under the assumptions of the Interpersonal Addition Theorem, it follows that if the Principle of Personal Good holds, then the general utility of an outcome can be written as the sum of individual utilities. And if the general utility of an outcome can be written as the sum of individual utilities, then the Principle of Personal Good holds. Hence, we can trust the Principle of Personal Good *if, and only if, the general utility of an outcome can be written as a sum of individual utilities*. This is quite a strong condition for the aggregation of individual utilities.

In the case of valuing life, it follows that the ex ante willingness of all individuals to accept the increased risk of death is a reason to accept the project only if, in each state, the sum of the individual utilities of benefits and compensation is larger than the (dis)utility of the death of one individual. In other words, it is not only necessary, for all persons, that the  $(n-1)/n$  chance of receiving benefits and compensation outweighs the  $1/n$  chance of death; it is also necessary that benefits and compensations for  $(n-1)$  individuals outweighs the actual death of one individual. Again, this is not to say that the Principle of Personal Good cannot be justified; but it clearly illustrates that strong premises are needed.

## Equality in Outcomes

It follows directly that the Principle of Personal Good conflicts with egalitarianism, as it is commonly understood, namely, that inequality in the distribution of good among individuals has negative value. John Broome (1991b, p. 185) has introduced the following (☞ [Table 16.10](#)).

Interpret the figures as standing for people's good. The idea is to demonstrate a conflict between the Principle of Personal Good and egalitarianism. The two alternatives are equally good for both individuals (and this is so regardless of the attitude to risk, because the risk is the same for both in both alternatives). Hence, the Principle of Personal Good implies that they are equally good. However, an egalitarian would prefer A, because it is certain to lead to an equal outcome. So an egalitarian is forced to deny the Principle of Personal Good.

▣ **Table 16.10**  
**Equality under uncertainty**

	Act A		Act B	
	State heads $p = \frac{1}{2}$	State tails $p = \frac{1}{2}$	State heads $p = \frac{1}{2}$	State tails $p = \frac{1}{2}$
Individual P	2	1	2	1
Individual Q	2	1	1	2

However, there is the possibility that inequality could be considered a personal harm. If the negative value of inequality is part of an individual's good, there is no conflict with the Principle of Personal Good. Broome (1991b, pp. 192–196; see also Broome (1990–1991), reprinted in Broome (1999, pp. 111–122)) has argued that this maneuver is not ad hoc, but actually captures the essence of the concern for fairness. But this view has not been discussed in detail in the literature.

The Interpersonal Addition Theorem states aggregation across individuals in terms of utility. Harsanyi believed that his original theorem supported utilitarianism. But once we consider the complex relation between an individual's utility and his good, it is clear that the sum of individual utilities does not represent utilitarianism, as it is ordinarily understood. Rather, it is equivalent with some form of the priority view. I say "equivalent" for the following reason: If we assume that personal utility is an increasing strictly concave transform of personal good, this weighting function is determined by considerations of risk, and the priority view is concerned with giving more weight to those who are worse off for moral reasons. But for each form of risk weighting function, there will be an identical moral weighting function, such that the sum of personal utilities is equivalent to the version of the priority view having a moral weighting function of the same form.

## Further Research

---

Ethical theories about the distribution of good across individuals, such as egalitarianism, utilitarianism or the priority view, are typically stated in terms of goodness in situations with no uncertainty. As we have seen, decision theory seems to imply that the ordering of uncertain acts is more basic, because it is possible to take the value of avoiding risk to good into account. But then the ethical question becomes how to aggregate individual utilities in an outcome taking uncertainty into account, rather than goodness with no uncertainty. And this would appear to change the focus of ethical theory considerably.

I see three general directions for future research. One is to try to examine the relations of the two approaches by determining how an individual's utility relates to his good. A way to do this would be to model how goodness and considerations of risk each contribute to overall utility. Technically, this could be done by using what is known as a model for conjoint measurement, i.e., a form of measurement, in which the contribution of two or more factors to some overall magnitude can be measured simultaneously (Krantz et al. 1971, pp. 245–368). Bengt Hansson (1988) has hinted at this idea.

The other line is to simply abandon the traditional certainty approach and examine the implications of an approach inspired by decision theory in more detail. This raises questions such as: How should the concern for equality be stated in terms of individual utilities? Is there an important difference between utilitarianism and the priority view, if utilitarianism is stated in terms of individual utilities?

The third line is to abandon decision theory and to develop an alternative model for decision making under uncertainty and examine its implications for our understanding of individual and general betterness. For instance, it would be very interesting to have a model for uncertainty which does not build on the Sure-thing Principle, and then examine its implication for the ethics of uncertain decisions.

In all cases, there is a lot of work to be done.

## References

---

- Allais M (1953) *Fondements d'une Théorie Positive des Choix comportant un Risque et Critique des Postulats et Axiomes de l'Ecole Americaine*. Econométrie, Colloques Internationaux du Centre National de la Recherche Scientifique, XL: 257–332
- Allais M (1979) The foundations of a positive theory of choice involving risk and a criticism of the postulates and axioms of the American School. In: Allais M, Hagen O (eds) *Expected utility hypothesis and the Allais paradox*. Reidel, Dordrecht, pp 27–145
- Arrow KJ (ed) (1970) The theory of risk aversion. Essays in the theory of risk-bearing. Amsterdam, North-Holland, pp 90–109
- Bentham J (1789/1948) *The principles of morals and legislation*. Hafner, New York
- Bernoulli D (1738) *Specimen theoriae novae de mensura sortis. Commentarii Acad Scientiarum Imperialis Petropolitanae* 5:175–192
- Bernoulli D (1954) Exposition of a new theory on the measurement of risk. (trans: Louise Sommer). *Econometrica* 22:23–36
- Broome J (1978) Trying to value a life. *J Public Econ* 9:91–100
- Broome J (1990–1991) Fairness. *Proc Aristotelian Soc* 91:87–102
- Broome J (1991a) Utility. *Econ Philos* 7:1–12
- Broome J (1991b) Weighing goods. Equality, uncertainty and time. Blackwell, Oxford
- Broome J (1999) Ethics out of economics. Cambridge University Press, Cambridge
- Cantor G (1895) Beiträge zur Begründung der transfiniten Mengenlehre. *Math Ann* 46:481–512
- Davidson D, McKinsey JC, Suppes P (1955) Outlines of a formal theory of theory of value, I. *Philos Sci* 22:140–160
- de Finetti B (1937) La prévision: ses lois logiques, ses sources subjectives. *Ann Inst Henri Poincaré* 7:1–68
- de Finetti B (1964) Foresight: its logical laws, its subjective sources. In: Kyburg Jr. HE, Smokler HE (eds) *Studies in subjective probability* (trans: Henry E. Kyburg, Jr.). Wiley, New York, pp 93–158
- Debreu G (1959) Theory of value. An axiomatic analysis of economic equilibrium. Yale University Press, New Haven
- Diamond PA (1967) Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility: comment. *J Polit Econ* 75:765–766
- Ellingsen T (1994) Cardinal utility: a history of hediontometry. In: Allais M, Hagen O (eds) *Cardinalism. A fundamental approach*. Dordrecht, Kluwer, pp 106–165
- Fishburn PC (1970) Utility theory for decision making. Wiley, New York
- Gärdenfors P, Sahlin N-E (1988) Introduction: Bayesian decision theory – foundations and problems. In: Gärdenfors P, Sahlin N-E (eds) *Decision, probability, and utility*. Cambridge University Press, Cambridge, pp 1–15
- Hansson B (1988) Risk aversion as a problem of conjoint measurement. In: Gärdenfors P, Sahlin N-E (eds) *Decision, probability, and utility*. Cambridge University Press, Cambridge, pp 136–158
- Harsanyi JC (1955) Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *J Polit Econ* LXIII:309–321
- Harsanyi JC (1976) Essays on ethics, social behaviour, and scientific explanation. Reidel, Dordrecht
- Hicks J, Allen RG (1934) A reconsideration of the theory of value. *Economica* 1:52–76, 196–219
- Jackson F (1991) Decision theoretic consequentialism and the nearest and dearest objection. *Ethics* 101(3):461–482
- Jeffrey RC (1983) *The logic of decision*, 2nd edn. Chicago University Press, Chicago
- Jensen KK (1995) Measuring the size of a benefit and its moral weight. On the significance of John Broome's "Interpersonal Addition Theorem". *Theoria* LXI (Part 1):25–60
- Jevons WS (1873/1970) *The theory of political economy*. Penguin, Hammondsworth
- Kemeny JG (1955) Fair bets and inductive probabilities. *J Symbolic Log* 20(3):263–273
- Krantz DH, Luce RD, Suppes P, Tversky A (1971) Foundations of measurement. Volume 1: Additive and polynominal representations. Academic Press, San Diego
- Lehman RS (1955) On confirmation and rational betting. *J Symbolic Log* 20(3):251–262
- Luce RD, Raiffa H (1957) Games and decisions. Introduction and critical survey. Dover, New York
- Machina MJ (1982) 'Expected utility' analysis without the independence axiom. *Econometrica* 50:277–323
- Mellor DH (2005) Probability. A philosophical introduction. Routledge, London
- Moore GE (1903/1993) *Principia ethica* (revised edn). Cambridge University Press, Cambridge
- Page AN (ed) (1968) *Utility theory. A book of readings*. Wiley, New York
- Parfit D (1984) Reasons and persons. Clarendon, Oxford
- Peterson M (2004) From outcomes to acts: A non-standard axiomatization of the expected utility principle. *J Philos Log* 33:361–378

- Peterson M (2009) An introduction to decision theory. Cambridge University Press, Cambridge
- Rabinowicz W (1995) To have one's cake and eat it, too: Sequential choice and expected utility violations. *J Philos* 92:586–620
- Ramsey FP (1931) Truth and probability. In: Ramsey FP (ed) The foundations of mathematics and other logical essays. Kegan Paul, London, pp 156–198
- Robbins L (1935) An essay on the nature and significance of economic science, 2nd edn. Macmillan, London
- Savage LJ (1972) The foundations of statistics, 2nd edn. Dover, New York
- Schick F (1986) Dutch bookies and money pumps. *J Philos* 83:112–119
- Shimony A (1955) Coherence and the axioms of confirmation. *J Symbolic Log* 20(1):1–28
- Vallentyne P (1993) The connection between prudential and moral goodness. *J Soc Philos* 24:105–128
- von Neumann J, Morgenstern O (1944/2004) Theory of games and economic behaviour (Sixteenth Anniversary Edition). Princeton University Press
- Wald A (1950) Statistical decision functions. Wiley, New York
- Zimmerman MJ (2008) Living with uncertainty. Cambridge University Press, Cambridge



# 17 The Mismeasure of Risk

Peter R. Taylor

University of Oxford, Old Indian Institute, Oxford, UK

<i>Introduction</i> .....	443
<i>Risk</i> .....	446
<i>Chance</i> .....	447
<i>Harm</i> .....	449
<i>Time</i> .....	449
<i>Uncertainty in Outcomes</i> .....	450
<i>Thresholds and the Relativity of Risk</i> .....	451
<i>The Interrelation of Risk Metrics, Behavior, and Risk</i> .....	452
<i>Risk and Reward</i> .....	453
<i>Judgment</i> .....	454
<i>The Structure of Risk Decisions</i> .....	455
<i>Recent Illustrations of Risk</i> .....	456
Volcanic Ash .....	456
Hurricanes .....	457
Terrorism Insurance .....	458
The Banking Crisis .....	459
What Have We Learned So Far? .....	459
<i>The Measurement of Risk</i> .....	460
Unknown Unknown Risk .....	461
Model Risk .....	462
Measures and Thresholds .....	463
Risk and Reward .....	464
ROB .....	467

<i>Applications</i> .....	467
RAG .....	468
Solvency II .....	471
<i>Summary</i> .....	473
<i>Further Research</i> .....	474

**Abstract:** Tsunamis, volcanic ash clouds, financial crashes, and oil blowouts are recent disasters that have caught us by surprise and dominated world headlines. It seems that we are just not very good at predicting such risks or dealing with their consequences. This chapter looks at one reason this might be so – that we are not measuring risk appropriately, either in how we assess a risk or how we then make decisions about risk-related problems. Our measurements are conditioned by expectations of precision and simplicity which, we will argue, are too often lacking in the real world.

In assessing a risk we can miss potential outcomes, especially when a system changes over time. We can treat our theories as true irrespective of empirical support, and we can set impractical thresholds of acceptable risk. When we do measure the risk, we may not have enough data of sufficient quality, or behavioral responses may reduce the measured risk yet increase risk that is not measured. In short, our current assessments of risk are too blinkered.

The associated blunders in decision-making are also all too familiar – disregarding a risk because the model says the chance of the event occurring is low even though we have little confidence in the model; conversely, taking an unduly “precautionary” approach to avoid risks when the potential for harm is negligible; worst of all, perhaps, missing the wider decision-making context needed for a sound judgment.

Put together, these factors mean that simple measures of risk can be a poor guide for decision-makers. This chapter advocates five extensions to current methods in order to avoid these pitfalls in the future. In risk assessment, we need to:

1. Estimate *unknown unknown risk*
2. Quantify *model risk*
3. Use *multiple measures and thresholds*; and in decision support
4. Balance *risk and reward*; and
5. Examine *ROB* (Risk Outside the Box)

We will describe examples where, had certain measures of risk been in place, disasters might have been averted. We will then show how each of the extended methods may be applied. Two particular applications of the proposed approach will be provided – “RAG” statuses in project and risk management, and Solvency II in the insurance industry.

## Introduction

---

Worrying about risk has become a defining neurosis of our age (Beck 1992). The news is full of disasters and reports on what was not done in the past and recommendations about what needs to be done in the future. On the one hand, we are safer than ever before in our daily lives yet, on the other, our interdependent and technologically powerful society has increased the potential for major disasters. Controls cost money and constrain freedom, yet accepting known risks, however small, is increasingly politically unacceptable.

The purpose of this chapter is to show how we might plot a course through the conflicting demands of risk and control.

Many of the vagaries of life that were once just part and parcel of existence can now be anticipated and avoided or, at worst, managed if they do occur. Despite our efforts things still keep going seriously wrong, and, when they do, the subsequent investigation invariably concludes that warnings had been ignored and precautions not taken. Whether we are looking

at signaling systems on the railways, health hazards from asbestos, levees in New Orleans, the debt bubble from subprime lending, or deep water drilling for oil, we are wise after the event about how foolish we were before.

It is not just about miserly shortsightedness, though. Where we do invest effort and money, typically in complex computerized models, we often still get it wrong. “Useless Arithmetic” by Orin and Linda Pilkey (Pilkey and Pilkey-Jones 2006) is one of several recent accounts that gives examples of erroneous models, joined recently by accounts of the 2008 credit crisis such as “Fool’s Gold” by Gillian Tett (2009) which exposes the misjudgments in the banking sector and government.

Many reasons are given for such failures. Hubris, conspiracy, incompetence, greed; they all play a part, no doubt. This chapter will argue, however, that a deeper malaise underlies the mishandling of extreme risks – an inability to measure risk appropriately. We think we know more than we do; we believe we are safer than we really are; we seek oversimplification of a complex world. We would like risk to be binary – we either have it or we do not – whereas for most problems we have levels of knowledge and degrees of uncertainty. What we have not done is to feed the uncertainty back into the process. Instead, we prefer to believe our own publicity, a trap which has echoed down the centuries.

The philosopher David Hume touched on it while reflecting on the problem of induction (Hume 1748):

- ▶ The forming of general maxims from particular observation is a very nice operation; and nothing is more usual, from haste or a narrowness of mind, which sees not on all sides, than to commit mistakes in this particular.

The psychiatrist Henry Maudsley described the fallacy in the nineteenth century (Maudsley 1867):

- ▶ (we have) A sufficiently strong propensity not only to make divisions in knowledge where there are none in nature, and then to impose the divisions on nature, making the reality thus conformable to the idea, but to go further, and to convert the generalizations made from observation into positive entities, permitting for the further these artificial creations to tyrannise over the understanding.

From the twentieth century, the physicist Erwin Schrödinger (1935), reflecting on quantum mechanics, said:

- ▶ Reality resists imitation through a model.

and, coming right up to date, business is recognizing the danger of computer models as expressed by Tad Montross, CEO of General Re Corporation (Montross 2010):

- ▶ Understanding the models, particularly their limitations and sensitivity to assumptions, is the new task we face. Many of the banking and financial institution problems and failures of the past decade can be directly tied to model failure or overly optimistic judgements in the setting of assumptions or the parameterization of a model.

We may have lost the mechanistic certainties of traditional science, but our way of thinking is still primarily reductionist in that we presume we can explain the behavior of systems in terms of the behaviors of their component parts. Fundamental physics, founded on the precept of a mathematical theory that describes the way the world works, has been wondrously

successful. Yet its success has set in us an expectation that all problems can be reduced to equations from which we can compute answers to any problem. In practice, though, we have to use approximations and simplifications which we term *models*. Such simplifications can be known to be inaccurate or wrong yet the illusion of a fundamental theory remains which, supported by mathematical apparatus and computer calculations, makes it all too easy for us to believe that the model is the reality it claims to describe. Such *reification* can lead to serious misjudgments as most systems are too complex for fundamental theories to be used without simplification, approximation, or entirely remodeling. The inherent complexity of biological (e.g., Noble 2008) and economic (e.g., Beinhocker 2006) systems show that they are not adequately described by deterministic or traditional probabilistic models. Yet the siren call of models is seductive and too easily they take on the status of false gods.

The good news is that delusory false accuracy from the reification of models can in many important cases be addressed through explicitly recognizing *model risk*, which is the risk that the model is inappropriate to the problem. The bad news is that this is generally difficult and time-consuming, requires an element of subjective judgment, and meets resistance as it challenges preconceptions and vested interests.

A further twist is whether the various models allow for all potential outcomes. In one sense, of course, they never can and we end up assigning a *catch-all probability* to the set of unforeseen outcomes. The problem with this device is how to set the values for the catch-all probability and, even more so, how to estimate its development over time as circumstances change in our complex world. It is interesting to reflect on how well and poorly the future has been foreseen. For instance, leading American figures were polled in December 1900 ([http://www.yorktownhistory.org/homepages/1900\\_predictions.htm](http://www.yorktownhistory.org/homepages/1900_predictions.htm)) and asked to make predictions for the year 2000. In some cases, such as television, the predictions were remarkably prescient. Others, such as missing aeroplanes by concentrating on airships, reflected the technology of that time. Many, though, such as computers, nuclear power, and genetic engineering, were missed completely. With the current increased pace of technological change, we can anticipate the next 20 years to have as many changes as the last 100.

Recent disasters have reawakened awareness of unforeseen outcomes with Donald Rumsfeld's *unknown unknowns* (Rumsfeld 2002) and Nassim Taleb's *black swans* (Taleb 2007). *Unknown unknowns* are those outcomes we neither knew nor could have known "the ones we don't know we don't know" as Rumsfeld described them, and *black swans* are exceptional and extreme occurrences falling outside the set of expected outcomes based on the problem of induction, and illustrated by the inductive conjecture that all swans are white. Taleb, in particular, sees a false parallel, which he terms the *Ludic Fallacy*, between real-world risk with its unknown outcomes and games of chance with their fixed outcomes.

Even if we allow for as-yet unknown outcomes and we estimate model risk, the most important job remains outstanding – judging what constitutes *acceptable* risk. This is rarely a one-sided judgment, as decisions involve considerations of benefits and costs, social and political norms, and ethics. While an assessment of benefits and costs – or risk-reward – can be quantified, setting their levels of acceptability within society and embracing ethical considerations makes this a judgment with a rational basis and not just a computation.

The choice of measures and thresholds alone may not allow the risk to be adequately measured. For example, a regulator may set one measure of risk to protect the public, whereas a company may use quite another measure of risk to decide their appetite for risk in order to make profit. Both are measures of risk but for different purposes.

Still this is not the end of the story, as the very measurement of the risk can affect behavior which in turn affects the risk. The “law of unintended consequences” is familiar in many areas where the vague is made precise, and changes to improve performance against one measure can cause adverse effects elsewhere in the system. Where this is done deliberately to manipulate the system we term it “gaming.”

In summary, then, there are factors over and above those in place today which can lead to the mismeasure of risk: unknown unknowns, model risk, thresholds for acceptable levels of risk, trade-offs of risk against rewards, and the inadequacy of single measures. Finally, there is an element of judgment to reflect factors outside the models which may increase risk.

The title of this chapter echoes Stephen Jay Gould's 1985 book “The Mismeasure of Man” in which he argued that the use of a single number, the IQ, as a measure of man's capability was a misguided simplification. As with models of risk, Gould identified as culprit the fallacy of reification (Gould 1985):

- ▶ The spirit of Plato dies hard. We have been unable to escape the philosophical tradition that what we can see and measure in the world is merely the superficial and imperfect representation of an underlying reality.

We will show how this fallacy of reification identified by Gould applies also to measures of risk, with similar consequences of incorrect judgment and action. Note, though, that in the past 30 years many of the criticisms made by Gould concerning IQ were themselves questioned because it turned out that the single number of IQ, when appropriately measured without cultural bias, was, after all, an accurate predictor of general intelligence. So it might well be with what one might see as the financial risk counterpart of IQ once we understand how to balance its measurement with other assessments. Just as IQ measures intelligence relative to a particular set of tests, so a measure called *VaR* (*Value at Risk*) measures financial risk relative to a particular model. VaR is currently under criticism following the banking crisis, yet may turn out to be a sound measure.

The main body of the chapter is divided into four sections. The first explains what we mean by *Risk*, the second works through some illustrative examples, the third sets out a constructive proposal on how to measure the risk of infrequent extreme events, and the fourth looks at two applications of the proposal – “RAG” statuses used in project and risk management, and the measure of risk used for Solvency II in the insurance industry within the EU.

## Risk

---

Risk applies to most aspects of our lives. We talk about the risks of rain tomorrow, of breaking our leg when playing a football match, of the collapse of a bank, of losing an election, of damage to our children's health from watching too much television, of an earthquake, or of a core meltdown in a nuclear power plant. We have financial markets dedicated to calculating and profiting from risk. In society we seem obsessed with attempting to eliminate risk through health and safety regulations, laws against smoking, and the seeming omnipresence of CCTV.

But what might such uses of the word *risk* have in common?

They share a feeling of uncertainty and harm – there might be some (possibly very severe) harm caused by our actions or nonactions or by certain events, but whether it will occur or how severe it may be, we do not precisely know.

At first glance, uncertain outcomes seem to call for a mathematical treatment in terms of probability theory. Indeed, the simplest mathematical notion of risk is “Probability times Harm” which is often still taken as a rational standard for risk assessment. Authorities like the Royal Society promote this notion of risk (Royal Society 1992) as:

- ▶ The probability that a particular adverse event occurs during a stated period of time, or results from a particular challenge

This sounds seductively like just one number, the probability, but in most cases this simple concept is not able adequately to capture the complexity of real-world risks. To the question “What is the risk of X?” we cannot sensibly expect to get back a single number as an answer unless X is a binary outcome. More generally, we should expect “there is so-and-so chance of this, and such-and-such chance of that, and . . .”; that is, there are multiple chance/harm potentialities. As these chance/harm combinations refer to some period of time, we end up describing risk in (at least) a three-dimensional risk space – *chance, harm, and time*. Decisions about risk will extend this definition further to a five-dimensional risk-decision space – *chance, harm, time, reward, and judgment*.

## Chance

---

Theories of chance have a long history of practical success yet interpretational difficulty. Chance can be about something that might happen (termed *aleatory*) such as “the chance of throwing a six is one sixth.” It can also be about what has happened but we do not know the answer (termed *epistemic*) such as “the chance of a six for a die that has already been thrown, but the outcome of which I do not yet know, is one sixth.” Chance and probability are largely interchangeable in normal usage, though in this chapter the term probability will be used to describe both the theory of chances of possible outcomes as well as the chance of individual outcomes themselves.

Put simplistically, there are two schools of thought about the meaning of chance. Some, termed *Frequentists*, view chance in terms of the statistical outcome of repeated actions. Others, termed *Bayesians*, think of chance in terms of our beliefs about the possible outcomes. Generally, we will view chance through Bayesian eyes as statements of belief, whether about a prediction or an as-yet unknown fact. The choice does not matter in practice as we will argue below that all our probabilistic statements are relative to models of reality, but it is easier to fall into the *reification* trap of thinking the model statements actually are the reality with the Frequentist view than it is with the explicitly subjective Bayesian view.

Whatever our interpretation, probability in its mathematical sense is a mapping of outcomes to numbers between 0 and 1, or in terms of percentages, between 0% and 100%. For instance, the outcomes of a fair die are the faces appearing up with values 1 through 6, and each with a chance of one-sixth. It also has some rules of additivity so that the chance of an actual outcome within a set of possible outcomes is the sum of chances of the discrete outcomes. For example, the chance of throwing a die value greater than 3 is 50% (3 out of 6). This sounds reasonable for a game of chance like dice, but the specification of individual probabilities seems overly precise for more general circumstances such as the chance of getting run over on the way home. In these scenarios, it is more natural to say we think there’s a so-and-so chance but with a such-and-such level of confidence.

For instance, we might say based on statistics of road accidents for those traveling home from Oxford that we reckon that the chance of getting run over tonight, all things being equal, is between 1 in a million and one in ten million. We might then say that our confidence that all things will be equal is only 50% as there is a major project to complete which requires late working. We can now envisage another scenario – having to stay in the office overnight. The chance of getting run over tonight if staying at the office is zero, so given the odds of staying in the office are 50% then the effective chance of getting run over tonight is reduced to between one in 2 million and one in 20 million. Another way of looking at this is that the first set of outcomes ignored the case of staying in the office. By including the chance of staying in the office, and assigning a chance to that higher-level outcome, we recover the probability of all outcomes even though they are now different from the “all things being equal” scenario. The advantage of the two-level response to the question shows that we have based and can explain our calculation on two levels of thinking: the chance of an outcome given that a model is assumed true, and the chance that the model itself is valid.

So *chance* really seems to mean both *primary probabilities* – “here’s a set of outcomes and their probabilities for a particular view” – and *secondary probabilities* – “here are the probabilities of the various views being right.” We can then combine these together to yield a net set of *effective probabilities* or, simply, *probabilities*!

Looking at this problem more generally, the calculus of probability assigns, in the technical sense of a *measure*, a chance to any collection of outcomes. It is not so clear how probability theory deals with ranges of precision about its assignments of chance. The range might be expressed as a range of uncertainties about the outcomes, or as ranges of uncertainties about the chance of given outcomes, or, more generally, as the probability distribution of sets of outcomes (Kaplan and Garrick 1981). Either way, we can always reexpress the outcome as another – *effective* – probability measure over the space of outcomes. But hold on, you might say, surely the second-level probability itself is subject to uncertainty? Should not there be a third-level probability and fourth-level probability and so on ad infinitum? Indeed, there should, and this infinite regress (see Atkinson and Peijnenburg 2006) bedevils the theory of probability. We do not propose any resolution to it in this chapter other than to say we can terminate the hierarchy with a statement of belief (which is essentially the solution offered by Kaplan and Garrick). What is material, then, is that the expression of the problem in terms of both first-level *and* second-level probabilities has more information and may thereby be more persuasive than simply the effective probability distribution.

Taking another example, we may think it highly likely that Everton will beat Tranmere Rovers in the football match tomorrow. They are local rivals, which can be expected to heighten motivation. The match might not happen, the Everton team might all have a bad day, Tranmere Rovers might play out of their skins; yet, despite these uncertainties, we still make statements about the chance of outcomes. We also make such statements in formal risk assessments where we use qualitative categories of likelihood and adversity – there’s a “HIGH” chance of some (harmful) outcome.

It may seem from this that “chance” cannot be expressed as probabilities after all. Yet, even in the case of the football match, we can get some numbers. After all, we might like to bet on the outcome and need some odds. It is not that odds cannot be set, as this regularly happens at visits to the bookmaker. The question should be rather, “how can this be done?” Well, implicitly in this case, explicitly in others, we create a model of the way football teams’ results depend on variables such as team member skills, recent form, condition of the pitch, position in the league table,

opinion of the pundits, content of the respective managers' mothers' dustbins for all we know. On the basis of that model, we can then calculate the risk. We could create multiple models. We could create many versions or parameterizations of any one of the models. Each would make different predictions. If we can test the models against some real-life data, we can start to gain confidence in adjusted combinations of models and their parameterizations. So the message here is that to construct chances, we necessarily employ a model or models of the situation.

## Harm

---

What constitutes harm? Harms might not be easy to compare. Is a bad cold worth something? Is a broken leg worth more? Even when we leave aside the interpersonal comparability between the harm as experienced by different people, it seems odd to assign some value to a broken leg, and another value to a cold. We do not as a matter of course say "I would rate a cold with 3, while a broken leg is worth 14 points on some scale that quantifies harm," rather we say: "I prefer having a cold to breaking my leg."

However, in order to be able to use quantitative risk tools, it does not suffice to have some ordered relationship between various harms; we need to assign actual values to them. For financial services, the simplest expression is monetary value and insurance, for instance, does put values on compensation for injuries, even going as far as government-approved tables such as the Ogden tables (see, for example, The Stationery Office 2007).

However callous it may seem, financial guidelines on the value of life and injuries are commonly used in economic assessments, with bodies such as the Judicial Services Board providing guidelines for personal injury insurance awards (JSB 2010).

While the most commonly used metric for harm is monetary value equivalent, mortality is also used, as in the Danish government's assessment of acceptable risk described below, and proxies for harm such as tons of emitted greenhouse gases are used to set policies on climate change.

As with estimates of chance, harms come with their own uncertainties and, as with probabilities, have second-order estimates of confidence in their values. These can be represented by harms having their own probability distributions, as in the range of damage to a specific building type for a given magnitude of catastrophic event. As with secondary probabilities, these estimates of confidence can be folded into the primary probability distribution to yield an effective probability distribution.

## Time

---

The importance of the time dimension to risk is often overlooked, yet is crucial to any validation or assessment of consequences. The timescales can influence which type of model might make the most robust predictions, and even whether it is possible to model the risk in any quantitative sense able to give predictive results. This is true both for the chance and the harm components of risk.

For short periods of time, there is often little change in either the range of possible outcomes or in our definition of what constitutes harm. Even better, over short periods of time we have become accustomed to being able to make predictions based on models. Indeed, we have grown so pleased with ourselves that we almost believe the models are the reality or believe they approach a "perfect model." Neither, though, seems to be justified in most practical assessments of risk.

Time is often overlooked in risk assessments on the grounds presumably that the world is fixed. We have seen, though, that time is one of the primary causes of *unknown unknown* uncertainty. There is one discipline where the time dependence of outcomes is of primary concern and that is *project management*. In project management one is balancing the factors of scope, resources (and cost), and timelines. One crucial factor in project management is assessing the chances of completing steps, tasks, projects, and programs on time. This is, perhaps, the most important part of project management as scope shifts and resource burns can often be a function of running over schedule.

## Uncertainty in Outcomes

---

The outcomes and their expected harm can be highly uncertain. The time factor magnifies the problem of uncertainty as probabilities can – and in most cases do – change over time; future generations, whom we might put at risk with today's actions or events, may have other preferences and experience harm differently.

How can we know there is not an outcome or state of a system we have not anticipated? And even if we think we know all possible outcomes, if it is a really complex problem, how can we have any idea at all of the chances of these outcomes? With many real-life problems being so vague, a consensus grew up that there are cases where uncertainty is different from probability (or risk). Keynes stated this explicitly in his theory of probability (Keynes 1937):

- ▶ The game of roulette is not subject . . . to uncertainty. The sense in which I am using the term [uncertain] is that in which . . . there is no scientific basis on which to form a calculable probability whatever.

Knight (1921) also famously defined risk as randomness that could be quantified as probability, and uncertainty, now often termed Knightian Uncertainty, as that which could not. Shackle (Shackle 1961) similarly pronounced that:

- ▶ Decision is not choice amongst the delimited and prescribed moves in a game with fixed rules and a known list of outcomes of any move or sequence of moves.

Nassim Taleb takes a particularly militant line against Knightian Uncertainty (Taleb 2006):

- ▶ In real life you do not know the odds, you need to discover them, and the sources of uncertainty are not defined. Economists . . . draw an artificial distinction between Knightian risks (which you can compute) and Knightian uncertainty (which you cannot compute) after one Frank Knight . . . Had he taken economic or financial risks he would have realised that these “computable” risks are largely absent from real life! They are mostly laboratory contraptions

There's certainly something odd going on with probability when it comes to decision-making under uncertainty. The *Ellsberg Paradox* (Ellsberg 1961) offers a compelling illustration of our preference for epistemic over aleatory uncertainty which might also be thought of as a preference for probability over uncertainty. (In a simple version of the Ellsberg paradox there is an urn with 90 balls, 30 of which are red and the other 60 may be either yellow or black, we do not know the mix. The question is would you prefer to accept the offer of \$100 on a random pick being a red or \$100 on it being a black? Most people choose the first option.)

There are many explanations given for the Ellsberg paradox. One such is that we make an overriding qualitative *judgment* about the risk based on our belief in the fairness of the offer, and in the case of the already chosen ball we subliminally anticipate the potential for fraud, as the asker could know the answer since the event has already happened.

Following Hubbard (Hubbard 2009) and Taleb, we shall not make the *Knightian uncertainty* distinction and have already argued that uncertainties can be subjectively quantified. However, when it comes to making a decision, we will propose that allowance of wider considerations – termed “Risks Outside the Box” or *ROB* factors – can adjust the importance assigned to a risk from purely quantitative considerations.

## Thresholds and the Relativity of Risk

---

In a recent talk at the Royal Society on the subject of uncertainty (Smith 2010), Lenny Smith announced “Solvency II is the gift of financial services to the physical sciences”! Could our derided financial systems actually be capable of teaching science something? He was talking about the importance of setting a threshold for unacceptable risk – in this case, the bar being set by regulators of insurance companies is that they must have sufficient capital to be 99.5% certain of paying claims within a 1-year period. This is as part of the Solvency II EU regulations for insurers which we discuss later in the chapter.

This amounts to saying that risk is relative to some standard of (un)acceptability and finding an appropriate measure and threshold is the most critical challenge in formulating a practical risk assessment.

Simple enough, then – all we have to do is set a threshold of unacceptable risk. Would we, for example, think it acceptable, had we the choice, to have people handling asbestos with a chance of asbestosis, or conducting experiments in Oxfordshire with a very small chance of a thermonuclear explosion, or developing new biological weapons with a chance of a harmful biological agent escaping? Sometimes the law enforces behaviors. For instance, it is not nowadays considered acceptable to drive a car without a seatbelt on, or ride a motor scooter without a helmet, or play cricket at school without a helmet. Only a few years ago smoking carriages were the norm on London tube trains, yet would now be treated as an affront to human rights. The science has not changed and there are no strong reasons to suppose the intrinsic risk has changed either. What has changed, instead, is our attitude to the acceptability of those actions that give rise to risk.

Another example is society’s attitude to cholesterol, where although the scientific evidence is equivocal (e.g., Kendrick 2008), beliefs in causal links between eggs and cholesterol and cholesterol in blood to heart disease are used to justify a wide range of diets and drug treatments.

It is clear from these examples that acceptability of risk is relative to society’s views and needs to be explicated in order for a risk to be a “risk” we bother with. And it may be that we set the threshold too low compared to the evidential harm – as was the case originally with asbestos – or too high – as was the case with the recent atmospheric volcanic ash – and when better evidence or methods of measurement or even just a change in social attitudes arrives, we change the threshold.

## The Interrelation of Risk Metrics, Behavior, and Risk

To illustrate how the choice of measure can affect behavior, consider the exposure that insurers have to catastrophe losses. Regulators such as Lloyd's of London, and rating agencies such as Best's, use catastrophe scenarios to probe the sufficiency of capital of an insurer. These are sometimes called *Realistic Disaster Scenarios*. Let's consider one such scenario – a US Florida hurricane making landfall in Miami and causing an insured damage of the order of \$100 billion. The existence of the scenario test can then influence insurers to take on individual risks outside the track of the chosen hurricane event. Whether deliberate or unintentional, the result is a low estimated loss for the test scenario, but high risk to the untested hurricane scenarios with different tracks. The risk to the (irresponsible) insurer's balance sheet is therefore understated by the test.

One response to the potential mismeasure of risk is to repeat the process for many conceivable scenarios, with a chance being assigned to each scenario. This allows the estimation of the probability of an insurer experiencing a certain loss or higher and is termed *probabilistic loss modeling*. Commercial suppliers such as RMS, AIR, and EQECAT produce models to assist underwriters and regulators to assess risk in this way.

From such a range of losses and chances, a simple probabilistic criterion can then be set to measure the amount of money associated to a given chance of loss in any 1 year – the amount or more that could be lost for a stated percentage of years. This measure is called *Value at Risk* or *VaR* and is used throughout finance. For insurers under the new Solvency II regime, the chance will be set to 99.5% in a year, so that the amount of money represents the threshold of what could be lost 1 year in 200.

*VaR* has been criticized in recent years as a blunt instrument made even less effective when operating in a regime about which we have little knowledge and few proven models. The risk of mismeasuring risk due to a lack of scenario models has thus been transmuted into the risk that the probabilistic loss model misrepresents the expected losses for the little understood experience of 1 in 200 year losses. To confound this new measure of risk further, the extreme *VaR* choice is not a measure that underwriters would use for trading which would instead be a view of profit and loss at a timescale of the order of a few tens rather than hundreds of years. This then creates a further conflict as there are multiple measures of risk relating to the differing objectives of the regulator (to ensure sufficient capital to pay policyholders or investors) and the business (to maximize profit).

There are further views possible on behavioral responses to risks. For example, Wilde ((Wilde 1994), see also (Adams 1995)) argues for *risk homeostasis* where risk is assumed in line with the benefits derived, so that the forces of risk of harm and potentiality of benefit will in most cases come to a dynamic equilibrium (*homeostasis*) somewhat akin to a regulator or thermostat. In this way, minimizing a risk in one area displaces the risk to another.

Another, more sinister, take on behavioral response is *moral hazard* whereby one action that attempts to reduce risk can so insulate a party from the direct consequences of that risk that they increase their risky behavior. An example is where drivers become more reckless, or make an expensive choice of garage to repair damage after an accident, because they are protected from the financial consequences by insurance.

## Risk and Reward

In financial circles, *risk* is expressed in terms as financial loss. If loss relates to risk, then what does profit relate to? *Positive risk* of course! This is shown in Fig. 17.1 below where the negative x axis represents loss (or negative profit) and the positive x axis profit (or negative loss).

What we see here explicitly is that we could look at the downside (the “risk”) but it would be foolish to do so commercially in the absence of considering the upside (the “profit”). In this example, the profit is marginal relative to its volatility, so one might well be concerned whether this would be a successful business.

It is worth commenting further on the shape of this curve, as it will vary significantly according to the problem. Fig. 17.1 illustrates a portfolio of risk rather than individual- or catastrophe-exposed accounts, which have a more singular and bounded form. It also does not reflect the most important aspect of managing risk, which is to find ways to offset the downsides. One of the primary methods of dealing with this is, of course, insurance (or reinsurance if you are an insurer). We will return to this point in response to risk.

This assessment of upside as well as downside is sometimes termed *Risk–Reward* or *Cost–Benefit*, and the tolerance for downside risk is termed the *Risk Appetite*. We will return to these concepts as they apply to finance later in the chapter. For now, though, our personal experiences confirm that decisions concerning risk are not just about acceptability levels, they are also about whether an action that incurs risk has benefits, and how we balance these against each other. For instance, is the damage caused to the environment from bovine flatulence outweighed by the benefit from humans eating beef – what is the risk appetite for beef (!)?

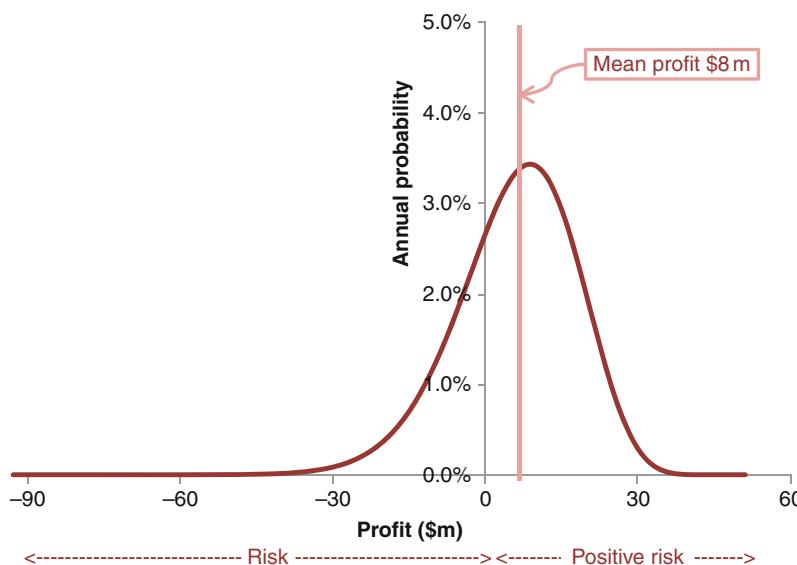


Fig. 17.1  
Positive risk

As well as the risks and benefits to an individual or corporation in relation to their actions, the uneven distribution of risk and reward often impels us to consider the wider consequences of our actions. This is particularly true of modern technological developments, many of which are *dual use* with benefit in one area or to one party having associated harm to another area or party. In such cases we have to consider the balance of consequences, and often these do not reduce to a utilitarian or financial measure from which a decision can be made. To give an example (POST 2009):

- ▶ Advances in DNA synthesis have enabled cheap and rapid synthesis of some viral genomes. There is concern that further advances and improved methods for delivering infectious agents may bring bioweapons production within the capabilities of terrorist groups.

Companies might unwittingly aid terrorists by providing DNA segments that could be joined to create the full genome of an infectious agent. In 2006, a *Guardian* reporter successfully ordered a small DNA segment of the smallpox genome for delivery to his home address.

The point here is that it is not just the *risk* that needs to be measured, it is the overall *risk-reward* or *cost-benefit* position, and this in turn will often come down to a judgment.

## Judgment

---

As Peter Bernstein said in the introduction to his book on risk (Bernstein 1996):

- ▶ The story that I have to tell is marked all the way through by a persistent tension between those who assert that the best decisions are based on quantification and numbers, determined by the patterns of the past, and those who base their decisions on more subjective degrees of belief about the uncertain future. This is a controversy that has never been resolved.

Given the complexity of the world and the inadequacy of many of our models to tackle risk, it is perhaps no coincidence that many of the risk assessments in business are qualitative, taking forms such as *Heat Maps* and *Risk Registers* in which risk is subjectively rated and mitigations identified. Some authors (e.g., Hubbard 2009) decry qualitative measures arguing that their predictive track record is poor. Indeed, we agree that many qualitative risk assessments can be lazy shorthands and that adoption of *Bayesian* subjective probability and recognition of secondary probability can sharpen risk assessment in many cases. However, there can still be factors that are outside the domain of the initial risk model for which judgment is needed. As an example, consider an insurance company which has, as its primary stated business exposure, loss due to a Florida hurricane. When asked off the record, though, the CEO might have loss of an underwriting team or withdrawal of a capital provider as her primary risks. The human factor is rarely far away from real-life risk yet rarely explicated in the risk model, albeit there is now a label for such risks – “operational risk.” Gerald Ratner, for example, famously put his company out of business with an injudicious comment about its products.

If judgment is an overlaying factor that emphasizes some risks beyond their strictly quantitative values – “risks outside the box” one might say – for chance and harm, what sort of measure is it? We propose that it is a categoric measure (such as the color-coding in *Heat Maps*) and represents the level of *importance* needed in decision-making.

## The Structure of Risk Decisions

We can now see five elements combining together in articulations of risk and decisions about risk – chance, harm, time, reward, and judgment. The schematic below illustrates these components and how they can project down to particular views of risk decision problems (Fig. 17.2):

The “projections” suppress or fix some of the dimensions, either because a dimension is irrelevant to the decision, or in order to make the problem tractable. For example, in evaluating projects, we often look at the chance of achieving an outcome – delivery on schedule – not necessarily the downside or upside consequences. Similarly, risk analysis tends to fix the time dimension and disregard the benefit dimension, while cost–benefit analyses fix the chance ranges (e.g., a “risk appetite”) and evaluate upside versus downside. In some cases, harm and reward might be on the same “dimension,” but generally they are quite different variables such as mortality and financial consequence. For instance, in a “dual use” technology, the harm might be mortality in the third world whereas the reward might be financial benefit to Wall Street investors.

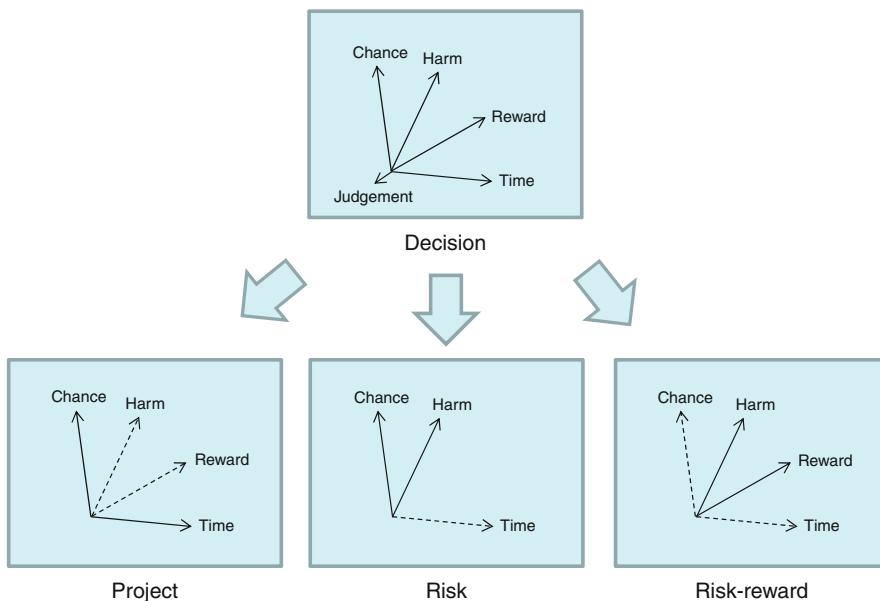


Fig. 17.2

The dimensions of risk decisions

## Recent Illustrations of Risk

### Volcanic Ash

The eruption of the Icelandic volcano Eyjafjallajökull in the spring of 2010 caused major disruption to travel and business. It came as a surprise to many, yet it had well-attested precedents. In 1783 Europe was plunged into a sulfuric acid smog resulting in mass death, summer heat, and winter freeze (Alfred 2010) from the Icelandic volcano Laki. It was also well known that volcanic clouds could cause problems after a passenger jet flew through the plume of Mount Galunggung in Indonesia on 24th June 1982. So although the threat was well known it was ignored (Cooke 2010):

- ▶ [OU's Dr Dave] Rothery has in his files a letter from British Airways dating from July 1989, turning down his request for funding for research into volcanic ash plume detection, on the grounds that it was "unlikely to be cost-effective".

"We've been saying there's a volcanic threat from Iceland for years and we need an ash monitoring system", says Rothery. "If we'd had a really good system in place, we'd have lost a couple of days flying and no more, I think. But the investment wasn't made".

In the absence of monitoring facilities, he says, data about the volcanic ash was based on computer modelling. "But you are only as good as your model. You need test flights, you need lidars – lasers looking upward from the ground, which give you a way of measuring the height and concentration and the particle size. And you can do the same from space not just with lasers but with infrared. There are all sorts of things you can do".

The same "head-in-the-sand" attitude applies to the thorny issue of how dangerous the ash actually is, and whether the six-day flight ban was justified he says. "There had been decades to work out what it was safe for an aircraft to fly through, and the airline industry hadn't done it. So the existing rule in place was: if you can detect any volcanic ash at all, you must not fly through it. That was just an arbitrary limit based on how good your detectors are".

Unsurprisingly, perhaps when it did happen, we were unprepared with untested computer models, no measurements of the effect of ash on air engines, and no acceptable threshold for flights to be suspended. As a result, we could not then avoid a precautionary overreaction with an economic loss estimated (Oxford Economics 2010) at \$5 billion and inconvenience to thousands of passengers whose flights were cancelled or delayed (The Economist 2010):

- ▶ No one denies that volcanic ash can cause jet engines to fail in flight. An engine's heat melts the fine-ground rock, which proceeds to encrust the cooler parts of the mechanism, stopping it from working. Lower concentrations can damage engines without having an immediate effect on how well they work. But where the boundaries between danger, potential damage and safety lie, and how they vary with the type and number of ash particles, was not taken into account in the decisions to close airspace. Things were made worse by the fact that the computer models of the ash cloud's dispersion gave only a very broad sense of where the ash might be.

By the early hours of April 15th, Britain's air-traffic controllers feared the worst. They had hoped the wind would carry the ash from Eyjafjallajökull south-east of the country, but the wind had changed. Guided by computer models used by volcanic-ash watchers at the Met Office, the air-traffic agency said it was withdrawing service. British airspace was closed.

Over the weekend both airlines and research agencies made test flights. Air France-KLM, British Airways, Lufthansa and others carried out over 40 flights. Subsequent engine inspection apparently revealed no unacceptable damage. On April 21st the CAA established a new rule, deeming regions thought to have less than 2,000 µg of dust/m<sup>3</sup> safe for flight. That threshold, the CAA says, was provided on the basis of data from equipment-manufacturers; Rolls-Royce, the leading European maker of jet engines for airliners, has made no comment on this. The new safety level is about 100 times higher than the background level of dust at ground level. It is also considerably higher than anything seen by research aircraft over Britain since the eruption started; those flights have encountered no patches of sky with an ash density of more than 400 µg/m<sup>3</sup>, 20 times the background level.

This example illustrates not only the importance of having measures and setting thresholds as we identified in the previous section, but that these are actually implemented.

## Hurricanes

---

Landfalling hurricanes are among the most damaging and costly events on earth. For the 2010 season which has just started, the conditions resemble those of 2005, the most active year on record which included hurricanes Katrina and Rita. The year 2009, by contrast, was benevolent like 2006, the quietest year on record. So what do we think is the risk of US hurricanes in 2010?

Is this a well-formed question? Checking our criteria, we have the *time* element (2010) so what are our measures of *harm* and *chance*? For *harm*, do we mean some physical proxy for harm potential like a certain windspeed at the eye wall or a storm surge in excess of a certain height? Or a harm that causes many deaths? Or one that causes a large economic loss? Or one that causes large loss to the insurance industry? What measure should we use? For the purpose of this discussion, let us measure harm by the insured loss due to property damage, as that is the basis for the index maintained by the Property Claims Services (PCS).

What, then, does *chance* mean? We could say we have a chance of having a certain magnitude of hurricane and this is associated to a given magnitude of insured loss. Or should that be the chance of a given level of loss or more? As we have discussed already, the answer is really a curve, with loss plotted against chance of exceeding a certain loss value. This curve is called the *EP* or *exceedance probability curve*.

How can we estimate the chance of a “severe” hurricane and what does “severe” mean, anyhow? Does it mean \$100,000 of insured damage or \$100 billion of insured damage? PCS assign catastrophe codes for events giving rise to \$25 m or more of insured property damage. An insurer could get losses from lots of different hurricanes depending on where they land, how much damage they cause, and how much risk they insure (the retained loss, the insured limits, and the share of the loss the insurer is taking). The way insurers have got round this variability is to estimate the *EP curve* using a computer model which simulates a wide range of events that damage the insured properties. The simulation could be on an *occurrence* (any single event in the year) or *aggregate* (total expected for all hurricanes in a year) basis.

We are still not done, though, because there’s a lot of uncertainty about the damage caused and, as already mentioned, *secondary uncertainty* (see AIR 2003 and RMS 2001) is applied to provide a revised *EP curve*.

There is a further complication as there are several models available and they can give quite different results for the same data and assumptions. The complication can be overcome by extending *secondary uncertainty* to give each model's results a different weight depending on their applicability to the risk. Choosing how to weight the various models is still a problem, though, and in many cases is a balanced judgment rather than necessarily being supported by objective evidence. For instance, one might have no particular view and so wish to sample from each model equally.

Yet we are still not finished. It is quite possible that we believe there is a systemic risk: for example, that the model used was miscalibrated or that this year is a La Nina year or that global warming has increased the frequency and intensity of storms (many scientists argue this, though many say we do not know, and some even argue the opposite). What do we do here? Well this is our *ROB* level of judgment outside the model. In some cases it is now possible to rerun the models with options such as *Near-Term Events Sets* which allow for a view on global warming. Ultimately, however, each person has to judge how to use the model results.

## Terrorism Insurance

---

Here we have a contrast between a “sophisticated” model as marketed by a leading provider of catastrophe modeling software and a simple pricing methodology adopted by some insurers (Shipley 2009). The sophisticated model uses game theory to analyze terrorist networks by modeling a wide range of factors (RMS 2010):

- ▶ The model quantifies the impact of all terrorist attack permutations on 3,400 high-risk potential target locations. Targeting patterns and prioritization emulates terrorist objectives and uses game theory to incorporate the effects of security and counter-terrorism measures. Appropriate attack modes can be selected to model the coverages and exclusions in place.

At the very least, such a model needs to set probabilities of attacks, probabilities of different choices of weapons, probabilities of different categories of targets, and estimates of the loss impact, which are highly dependent of course on the choice of weaponry and the selection of target (Woo 2004). None of these probabilities can be any better than an educated guess, as there is no serious numerical information to support them. In combination, these guesses produce something that has the appearance of a sophisticated mathematical model but a vast range of potentially credible outcomes.

An alternative approach is far simpler and makes no claim of statistical rigor. All it does is look at the aggregate insured values in various areas of the USA, and assume that terrorists will wish to make an impact so that major urban centers with concentrations of valuable assets are more at risk than suburban or rural settings. It then takes a guess at the maximum size of loss – can a conventional attack do more damage than a 9/11 given that there is and was no other similar concentration of property value anywhere in the world? Answer: unlikely – and the maximum number of such losses that the US government could tolerate over, say a 3-year period. All the insurer then has to do is charge a rate applied to insured values that will generate enough premiums to pay the selected amount of loss plus a profit margin.

This model is not very sophisticated, but it has no spurious accuracy and no false comfort!

## The Banking Crisis

---

The start of the new century saw a delusion that debt had been mastered by new financial derivatives which spread the risk objectively at a fair price as described, for example, in (Tett 2009) and (Lanchester 2010). Far from spreading the risk, though, these instruments concentrated and spiraled debt. Crucial to this spiraling of debt, other than the usual litany of greed and collusion, was a set of computerized models which appeared to vaporize the risk of default on mortgages and loans. Many commentators at the time (such as John Mauldin in his *Outside the Box* e-letters) pointed out that this was a massive folly but, as with bubbles of the past, the road back to reality required a massive crash. One of the primary causes of the crisis was mispricing of mortgage default (credit) risk using computer-based mathematical tools of convenience, such as Gaussian copulas (Salmon 2009), which incorrectly estimated the chance of a borrower defaulting and the correlation between defaulters. Our analysis in the previous section would identify this as a “risk of risk models” or *Model Risk* and, indeed, the Turner Report in 2009 identified several major weaknesses in regulation and behavior, most notably on the use of models (Turner 2009, p. 22):

- ▶ The very complexity of the mathematics used to measure and manage risk, moreover, made it increasingly difficult for top management and boards to assess and exercise judgement over the risks being taken. Mathematical sophistication ended up not containing risk, but providing false assurance that other *prima facie* indicators of increasing risk (e.g., rapid credit extension and balance sheet growth) could be safely ignored.

Modeling problems were, it says (Turner 2009, p. 44):

- ▶ Short observation periods
- Non-normal distributions
- Systemic versus idiosyncratic risk (Correlated behavior)
- Non-independence of future events (The past is not a guide to the future)

Since the collapse of Lehman’s and the crisis that followed, the scales have fallen from the eyes of those blinded by the credit bubble. Model risk has emerged as a crucial hole in the Basel II framework which defined risk management.

Yet it seems that not all the lessons have been learnt as Turner rounds off his report (Turner 2009, p. 45) with:

- ▶ But it would also suggest that no system of regulation could ever guard against all risks/uncertainties, and that there may be extreme circumstances in which the backup of risk socialization (e.g., of the sort of government intervention now being put in place) is the optimal and the only defence against system failure.

This seemingly sensible statement has a subtext of moral hazard that could be the Achilles heel of bank regulation. Do not worry, it implies, take the profits but, if it goes wrong, argue it is an “extreme circumstance” and let the government socialize the losses.

## What Have We Learned So Far?

---

The four illustrations used highlight the potential inadequacies of the measurement of risk. The volcanic ash fiasco highlights the need for a quantitative measure and threshold; the

hurricane example shows how a risk problem, even when articulated, has to deal with model risk and judgment; the terrorism insurance case shows how simple alternative models can help us understand risk in a practical way, and the banking crisis shows how extant risk measures failed to pick up massive systemic risk.

Clearly the current methods of risk assessment and management are seriously flawed. What can be done?

## The Measurement of Risk

---

The general considerations of risk and analysis of recent disasters indicate the changes which might make the definition and use of risk measurements more effective. This is not just a set of additional *processes*, but also a way of recognizing the *values* that society, companies, and individuals place on risk, whether retaining, assuming, or ceding the risk. Thus, while the process change might be to set an explicit threshold for unacceptable risk using practically achievable metrics, it is our values which determine the measures we choose (e.g., mortality or financial loss) and their levels of acceptability (e.g., 10 deaths per incident or no more than \$1 billion). We shall see that the recognition of values becomes even more important at the decision-making part of risk management, where judgments are made in contexts outside the domain of the risk problem.

For instance, the top risk identified by a quantitative risk analysis of an insurance business might be a severe hurricane hitting Miami this year, yet the risk uppermost in the mind of the CEO might be withdrawal of support of the company's capital provider due to investigation of a white collar crime. Here's a risk which is outside the risk management process due to its confidentiality and high level of uncertainty yet will necessarily govern the CEO's immediate actions in the business.

As another example, consider two projects in a program, one of which is of minor financial value and another which incurs large penalties if it fails to deliver. The former project is deemed to be status RED and the latter status GREEN. What is the status of the program? Is it GREEN? What if the first project is the €10,000 decoration of the conference room for the US President's visit to the EU to discuss waste management in 1 week's time whereas the second project is the delivery of a €10 m report on waste management report to the EU Council of Ministers in 3 month's time? Maybe RED would more adequately reflect the political importance of the first project.

Many books and articles and procedure manuals exist covering the theory and practice of risk management. Many include the central ideas of definition of risk as a combination of harm and probability, the use of models, and the choice of a measure (and maybe also a threshold). So we take these for granted to avoid repetition and instead address five areas of extension to current approaches:

1. Unknown unknown risk
2. Model risk
3. Multiple measures and thresholds

4. Risk and reward
5. ROB (there's always more Risk Outside the Box . . .)

We will describe each issue, then propose actions to address it.

## Unknown Unknown Risk

---

How can we possibly know what we do not know?

Some claim that the infrequent severe events that surprise us (beneficially as well as harmfully, but usually harmfully!) are *Unknown Unknowns* or *Black Swans*, and fatalistically argue that we cannot be expected to anticipate the un-anticipatable and should instead accept that some things are beyond our abilities to foresee. The logic seems unassailable – after all, we cannot know what we do not know, can we? Yet on closer inspection, these events usually are not such surprises after all, and we are anyhow adept at making allowances for the vagaries of events in our everyday lives. We may have forgotten that we do it and how it is done. First of all, the majority of surprise events were not unknown at all, it is just that we chose to ignore them – they were gray rather than black swans or “Unknown Knowns” in the sense that we knew about them but chose to disregard this knowledge. For example, was “9/11” a Black Swan? Hardly, as the possibility of an aircraft crashing into the World Trade Center was a known scenario and the possibility of both Towers going down had been considered by insurers (at least in the company for which I worked which insured the Port of New York Authority) albeit rejected as implausible. What was not considered was a deliberate attack that destroyed both buildings. Second, we can make allowances for *Unknown Unknowns* by looking at our track record of failure as practiced, for instance, in the insurance of difficult risks – it is an extra allowance in the premium for what we do not know about yet; a “margin of error.” Third, although the elements of unpredicted risk lurk unidentified in the present, generally they emerge as the risk landscape changes over time. The longer the period of time involved, the higher the risk of unforeseen events tripping us up.

Building on these observations, here are three suggestions on how we might identify *Unknown Unknowns*:

1. *Scenario tests.* By definition, these are a form of *known unknowns* so what scenario tests really do is consider responses to a challenging range of (generally adverse) possible futures in which (Wack 1985) “Decision scenarios describe different worlds, not just different outcomes in the same world.” Scenario planning is widely used in business, and quantitative examples in finance include *Lloyd’s Realistic Disaster Scenarios* in the insurance industry and *Stress Tests* and *Reverse Stress Tests* in both banking and insurance. The idea of a *Reverse Stress Test* is to ask what scenarios could breach capital limits, as opposed to a *Stress Test* which asks what the consequences of a particular scenario might be. (At the time of writing this chapter, though, banking regulators have not asked for scenarios of sovereign debt default, indicating they have some way to go in understanding the current world!)
2. *Past Experience.* One way to assess unforeseen outcomes is to check how well we have foreseen them in the past. If our track record is good, and the environment within which we are operating is not changing much, we might be justified in a low allowance for *Unknown Unknown* risk, but if the track record is poor it invites us to think through a greater range of options and models.

3. *Timescale.* Arguably, time is the greatest single enemy of risk management as the world changes and redefines the space of outcomes. The most obvious way to cater for this is through frequent reviews of risk, commensurate with the changes in the environment creating the risk. For instance, in the insurance business, quarterly risk reviews are carried out and especial care taken over multiyear policies. In a project, monthly or weekly progress reviews are often sufficient, but at critical stages it can be necessary to review project progress daily (in what is sometimes termed *daily prayers*) as there is so much risk to manage.

What might we do with better intelligence on *Unknown Unknowns*? We could take mitigative action – for example, reducing exposure or increasing prices for multiyear insurance policies if we are insurers. We could use the information to inform our choice of models and/or allowance for model risk. We might adjust our thresholds for acceptability of risk. Certainly, we could expect the assessment of *Unknown Unknowns* to color our view of risk appetite. Above all, though, it might inform our judgment of action relating to risk.

## Model Risk

---

Many authors have highlighted model risk as a primary contributory factor to the banking crisis – for example, the Turner Report (Turner 2009), reinsurance business executives (Montross 2010), and authors of handbooks (Gregoriou et al. 2010). In many cases, the choice of model and the variability consequent on that choice of model is the single biggest “risk of risks.” Overdependence on computerized models is a principal cause of erroneous decision-making especially where the call for simplicity causes a misstatement of the problem as, for instance, was seen so tragically in the overfishing of the Grand Banks (Pilkey and Pilkey-Jones 2006).

Despite this, regulators still shy clear of requesting explicit capital to reflect model risk. This may be to do with the psychology of requiring simplicity where it may not actually exist, or that the regulators feel able to factor in these uncertainties when setting the final regulatory capital number. A home has been found for “model risk” which has not been fed back into the quantitative risk assessment, and that is part of “operational risk.” Although there is variation of the definition of operational risk, it is generally (e.g., Basel 2010) held to cover “The risk of loss resulting from inadequate or failed internal processes, people and systems or from external events.” This includes the inadequacies of models. Whether in practice the model risk is adequately assessed and quantified, though, is quite another matter as in some cases the model risk at a “One in 200 Year” level would be substantial and there is little interest in management taking on additional capital requirement. Model risk is the Achilles heel of the regulators.

The problem seems clear, so what can be done? How can we ever have time to develop, examine and compare all the many models we can conceive for a given problem? As with *unknown unknowns* we seem to be caught in the trap of never being able to “know what we do not know.” Actually, we do already have alternative models. We even have methods of combining our degrees of confidence in such models. What we have not done is to incorporate such methods into risk assessments.

Here is a non-exhaustive list of checks that might be used to assess Model Risk:

1. *Model adequacy.* Is this model fit for purpose? There are various ways in which this might be tested. For instance: a careful evaluation of assumptions; a test for a particular scenario so that the consequences can be tracked through in detail rather than relying on the black box or

hoping the law of large numbers will somehow cancel out any systemic errors; comparison to other models; benchmarking against comparable risks. Such tests might be operated in concert. For instance, in probabilistic loss modeling in insurance, Model A might give a result quite different to Model B. One way to find out what is going on and which model to choose is to take some scenarios and compare them by individual component losses. We can then get a further view by taking the proportion of the estimated industry loss according to premium market share, and finally comparing (by class of business) the proportion of total insured value lost in each case against market benchmarks. These comparisons not only provide reasonableness tests but also flush out systemic errors. As we saw with the example of Terrorism Risk, for very complex problems it may be that computerized scenario-based models involve too many uncalibrated assumptions to be trustworthy, whereas a non-stochastic model can give a practical view of the risk based on simpler assumptions.

2. *Multi-model blending.* Supposing we have looked at the models and have decided that Model A and Model B are equally valid in our opinion, even though they give quite different results. If they are both scenario-based models it is possible to construct an effective selection of scenarios and associated losses by sampling from each set in equal measure. We would then end up with a new composite set of losses and thence a new composite risk curve reflecting our chosen blend. This blending can if required be made more sophisticated by allowing different blend mixes according to the type of scenario (for example, if we had greater confidence in Model A for large losses and Model B for smaller losses). The net result is nonetheless a composite event set.
3. *Recognition of model risk within a model.* This applies the idea of *Level 2 Risk* introduced by Kaplan and Garrick (Kaplan and Garrick 1981), which corresponds to the probability of probability models discussed above in the section  [Chance](#). It can also be applied in probabilistic scenario computations through the concept of *secondary uncertainty*. What this says is “Given that the event has occurred, what is the distribution of losses?” rather than “What is the mean loss for that event?” [*Primary uncertainty* in this terminology is the probability of the hypothetical scenario event.]

## Measures and Thresholds

Given a problem involving risk, we have seen that there are many challenges in deciding on measures of risk and thresholds of acceptable risk. While the general nature of the problem can be expressed in terms of risk curves of probability against harm, the metrics that are used to assess the risk can vary. Here are some examples of measures and thresholds:

- *Pollutants:* The concentration of the pollutant in air (or water), with maximum thresholds set according to their harmfulness. An example would include the various types of volcanic ash as they affect aircraft.
- *Mortality:* Although mortality is hardly acceptable to those directly associated with it, nonetheless levels of acceptability can be set in bands, as has been done by the Danish Government and the UK Health and Safety Executive (see Bedford and Cooke 2001, Chap. 18). Bands of fatalities (on a logarithmic scale) are plotted versus probability, in so-called *fC curves* (*fC* standing for frequency versus Consequence). The “acceptable” and “unacceptable” are then demarcated by regions of low frequency/low consequence and

high frequency/high consequence, respectively, with the middle region defined as “reduction desired.”

- *Value at Risk (VaR)*: As described above, VaR has vehement critics, with Nassim Taleb terming it “charlatanism” (Taleb 2006), and David Einhorn comparing it to “an airbag that works all the time, except when you have a car accident.” (Einhorn 2008). Their skepticism does not seem to imply that a probabilistic loss exceedance measure should not be used, but rather that the underlying models and the extreme value of the risk threshold is untestable. It could be reasonably responded that these are not in themselves flaws in VaR but rather in its application. Indeed, that is the view of this chapter subject to treating VaR as but *one measure* of risk, albeit a particularly valuable one when assessing downside consequences.
- *Credit risk*: In banking’s Basel II, regulatory capital requirement for credit risk offers a “standardized” measurement based on Ratings Agency ratings of the Counterparty. This incurs the further risk that the Rating Agencies get it wrong, as of course they did in the recent banking crisis.
- *The Learned Hand Formula* used in legal (tort) cases: This is especially interesting as it sets its own threshold to determine liability in an accident. It is an algebraic formula ( $B = PL$ ), according to which liability turns on the relation between investment in precaution ( $B$ ) and the product of the probability ( $P$ ) and magnitude ( $L$ ) of harm resulting from the accident. If  $PL$  exceeds  $B$ , then the defendant should be liable. If  $B$  equals or exceeds  $PL$ , then the defendant should not be held liable. (Grossman et al. 2006) explains how insurance markets collect and disseminate information about the expected values of all three variables in the Hand formula: the probability of accidents, the level of harm, and the burden of precaution.

## Risk and Reward

Decisions are rarely made only with regard to downside risk. In almost all cases, there are considerations of offsetting benefits or rewards. To take a current example, financial market regulators are requiring banks and insurers to hold levels of capital according to prescribed rules and processes as set out in the international Basel II accord for banks and, in the EU, Solvency II for insurance. But companies do not primarily want capital to protect against risk; they want it in order to generate a return for their investors. The risk part is an unavoidable necessity.

The diagrams below illustrate the typical probabilistic structure of profit and loss for two classes of business in an insurance company. The diagrams on the left show the probability (density) functions for profit and loss in a year for each class of business; the diagrams on the right show that the chance of losing \$30 m in either class is *the same*, and is 0.5% or once in 200 years, which is the Solvency II regulatory requirement for capital to support an insurance operation. Note that the two classes of business also have the *same mean profit* (☞ Fig. 17.3).

Although it is tricky to see, note the “fatter” tail of loss for Class of Business (CoB) 2. Other ways used to focus in on the purely risk perspective of these classes of business are shown in ☞ Fig. 17.4.

The *EP curve* view shows the *Exceedance Probability Curve* and, as you can see, the two classes of business cross over at the \$30 m loss and 0.5% probability level. What this means is that they have the *same chance* of 0.5% (1 in 200 years) of exceeding \$30 m loss in one year. The \$30 m is termed the *Value at Risk* or *VaR* for the 0.5% annual probability.

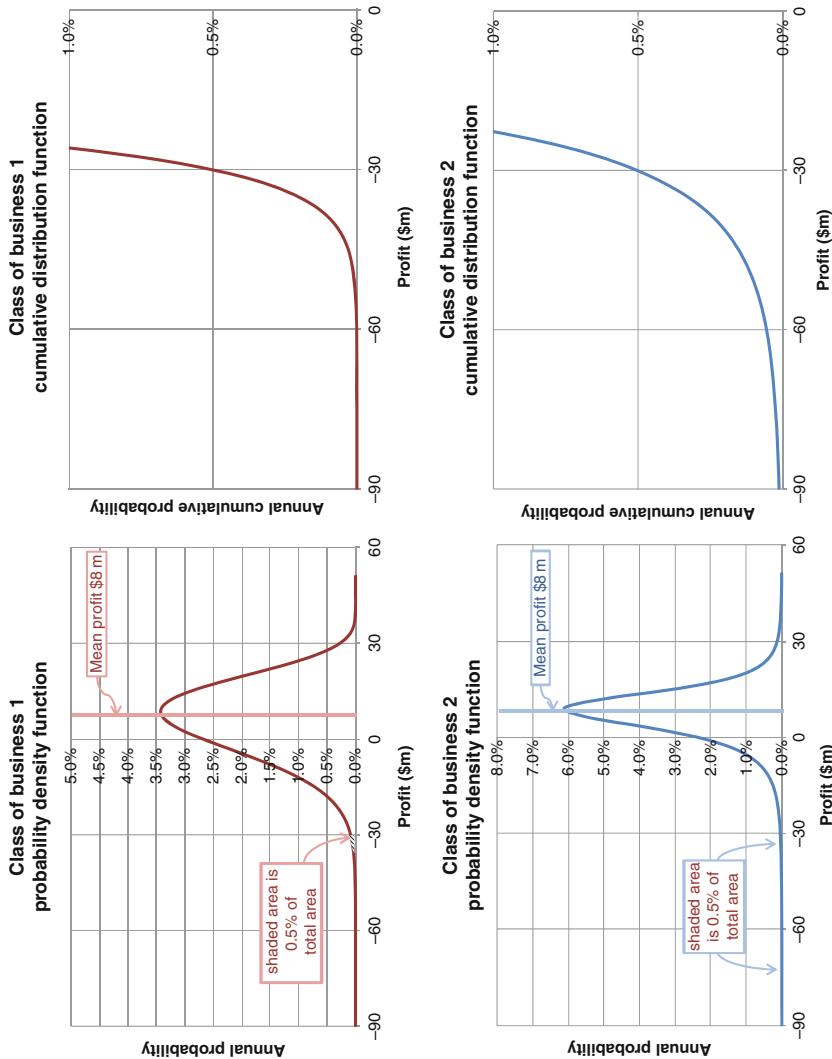
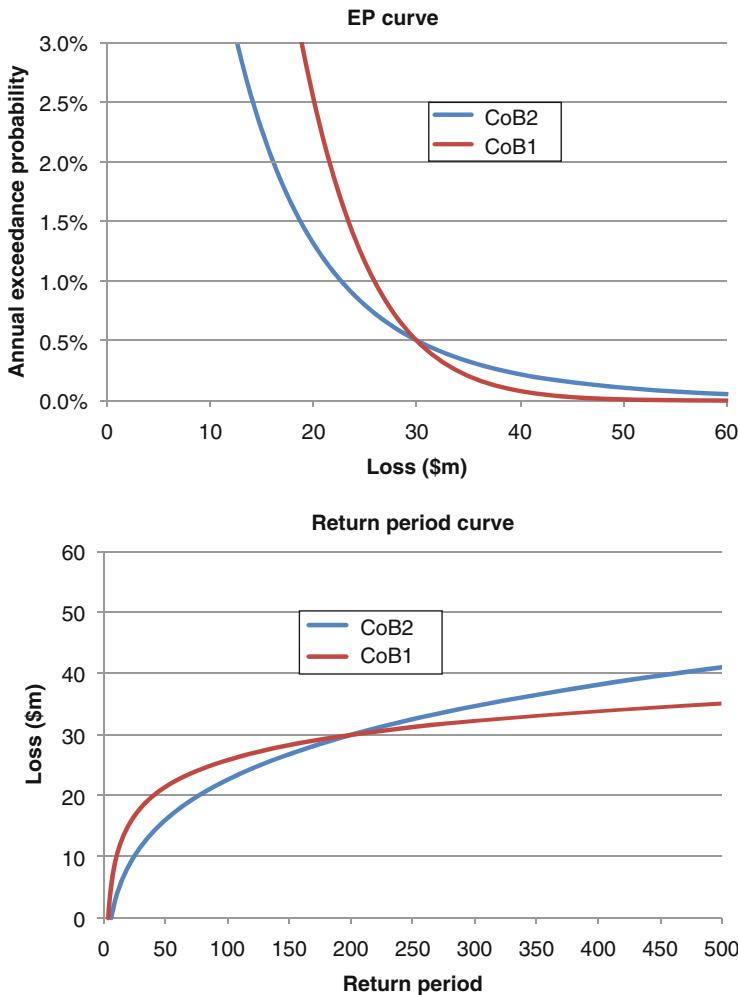


Fig. 17.3  
Two classes of business with the same 99.5% capital requirement and mean profit



**Fig. 17.4**  
Representation as exceedance probability (EP) curve and return period curve

CoB (Class of Business) 2 has a “fatter” tail (i.e., a larger chance of severe losses) than CoB 1 so it has a higher expected loss were it to breach the \$30 m VaR. This feature is often represented in a measure called the *Tail Value at Risk* or *TVaR* (also termed *Expected Shortfall*) which is the mean of losses *given that* the VaR loss point has been breached (for a given VaR probability).

In the *Return Period* view, the annual probability of exceedance is reciprocated to be expressed in terms of years (the *Return Period*) for a loss of that magnitude or more. Return Periods are often used in catastrophe insurance but are *not* intrinsic characteristics of individual catastrophes, being instead defined for a given insurance portfolio. What is then of interest is whether an insurance company’s Return Period for a given catastrophe is comparable to that of the entire industry. If it is not, then that insurance company is taking on risks with a profile different to that of the industry as a whole.

The Solvency II regulation sets capital based on the 200 year (99.5%) VaR point, but many insurance companies are practically assessing their risk at a lower Return Period of, say, 25 years using TVaR. This is because the lower Return Period represents risks they are likely to encounter in their business operation and TVaR allows for the extremity of the tail. Actuaries also like TVaR as a risk measure, as it satisfies the mathematical criterion of subadditivity (that the risk of the combined Portfolios A and B is less than or equal to the risk of A plus the risk of B) which VaR does not.

## ROB

---

Even having made the allowances described in the preceding sections, the world remains a risky place. There is always more Risk Outside the Box or *ROB*. Here's an example.

Suppose that Simple Insurance Ltd. has completed its Business Plan for the next year. The plan has taken into account stress test scenarios, reverse stress tests, multiple models, two different measures of risk and threshold, and optimized its portfolio based on an agreed risk appetite. All the risk boxes are ticked, the capital's in place, everything's fine. Or is it? The capital provider is rumored to be in trouble with subprime mortgages in Brazil. The commercial property underwriter is a maverick trader. The government is rumored to be putting a tax on all UK-originated business. Regulators are investigating questionable profit commissions. And so on and so forth. The environment in which we operate is so complex that we cannot necessarily represent all the risks we face in our models and measures.

*Risks Outside the Box* resemble *unknown unknowns* in that they look outside the extant model even though they are "known" by definition. We could spend our lives just worrying about all the many things that could go wrong, and even if we state them as assumptions (e.g., there will be no major change to regulation in the next year) there is still a risk that we might be mistaken. What can we do to identify *ROBs* and factor them into our plan?

In our Simple Insurance Ltd. example, the CEO, when asked his/her top risk, might answer "The Commercial Property Underwriter." Allowances for rogue underwriting might already exist in the measures of operational risk and some allowance might exist in the risk of failure of corporate governance. Yet how many companies are going to make significant allowance for a particular rogue trader as this is such a contentious and possibly legally difficult accusation? These types of risk slip through the quantitative models. The CEO can do plenty about this risk, though, as he/she might keep a close eye on the Commercial Property Underwriter's trading and develop contingency plans. The same applies to so many of the risks that affect us in life – we work at mitigating measures to bring them under control.

The most obvious action is to attempt to articulate the *ROB* (*See Risks Outside the Box*) back within the risk assessment process by incorporating it into the model, or making an allowance for model error, or maybe just overlaying a judgment category (e.g., flagging the risk as judged with RED importance). As the primary issue here is what decisions can be made based upon the risk assessment, *ROB* is primarily a matter for management.

## Applications

---

This section looks at two illustrations of the extensions to risk assessment introduced in the last section: the use of RAG (Red, Amber, Green) status setting, and the Solvency II regulations proposed for the insurance industry in the EU from December 2012.

## RAG

For those unaware of the term, *RAG Statuses* are categorizations of the state of something according to one of three values – Red (representing bad), Amber (representing neither bad nor good), and Green (representing good). That *something* might, for example, be a project, or a risk or a proposed action. In practice, more than three colors can be used, such as Purple for seriously bad, and Blue for future project tasks.

The gist of our analysis of RAGs is that we will think of them as *judgments* over and above an explicit quantitative rating of risk. The numbers may imply one RAG but there is the option to override it with judgment. This does not support the view that anything goes, rather that the difference between the analytic RAG status and the chosen RAG status is manifest and capable of rational justification.

RAGs are generally defined by ad hoc rules or set subjectively. In many applications they have an implicit meaning such as “chance of success” or “risk of failure” or “confidence in a beneficial outcome.” What we will do in this section is to flush out some of these meanings in a way that allows us to use RAGs in an analytical manner and, in particular, to explain *how to combine RAG Statuses*. Without a reliable method for combining RAG statuses, inconsistency can arise in the RAG statuses for phases of a project or risks within a class of risks. Note, though, that we always accept a RAG so constructed can be overridden to represent the *Judgment (ROB)* factor.

A popular application of RAGs is to project management. The idea here is that each project (or phase of a project or task within a phase of a project) receives a RAG status that represents the level of need for management attention. A project that is over budget and/or over time and/or failing to provide what it said it would is likely to be rated RED. A project on-time, on-budget, and delivering to scoped requirements is likely to be rated GREEN. Those in-between would be rated AMBER.

Applying our dimensions of risk, we might identify *Harm* as failure to deliver (perhaps with a financial consequence in contractual costs), *Reward* as meeting Requirements, and *Judgment* as the wider picture of whether the project has *Risks Outside the Box*, such as a reputational consequence even if the scope is small and the tangible costs and benefits are low.

In practice, many of the risk dimensions are fixed by the circumstances of a given project and the most important aspect of project management comes down to meeting deadlines, which involves the dimensions of *Chance* and *Time*. All sorts of factors can contribute to problems in meeting deadlines, such as lack of required resources (people, equipment), or failure of resources, or lack of money if over budget, or changes in requirements as problems turn out to more complicated than the original plan envisaged.

Taking a simple example, suppose we have two projects in a program of work, and that they are independent (albeit geared to a common objective). Modeling the chance of completion of a project by a Gamma function corresponding to a Poisson process, we can then use RAGs to represent ranges of the chance of completion in time.  *Figure 17.5* below shows how such bands could be set, how the probabilities of the individual projects change as the deadline is moved, and how the probability of the overall program changes on the basis that we are interested in measuring the probability of *both* projects being complete in time using the same banding.

You can see that we can have various combinations of component Project RAGs and an overall Program RAG that is more severe. The same probabilistic structure has been extended to networked Gantt charts akin to Bayesian Nets. The outcomes can then be computed by Monte Carlo simulation (e.g., Bedford and Cooke 2001, Chap. 15).

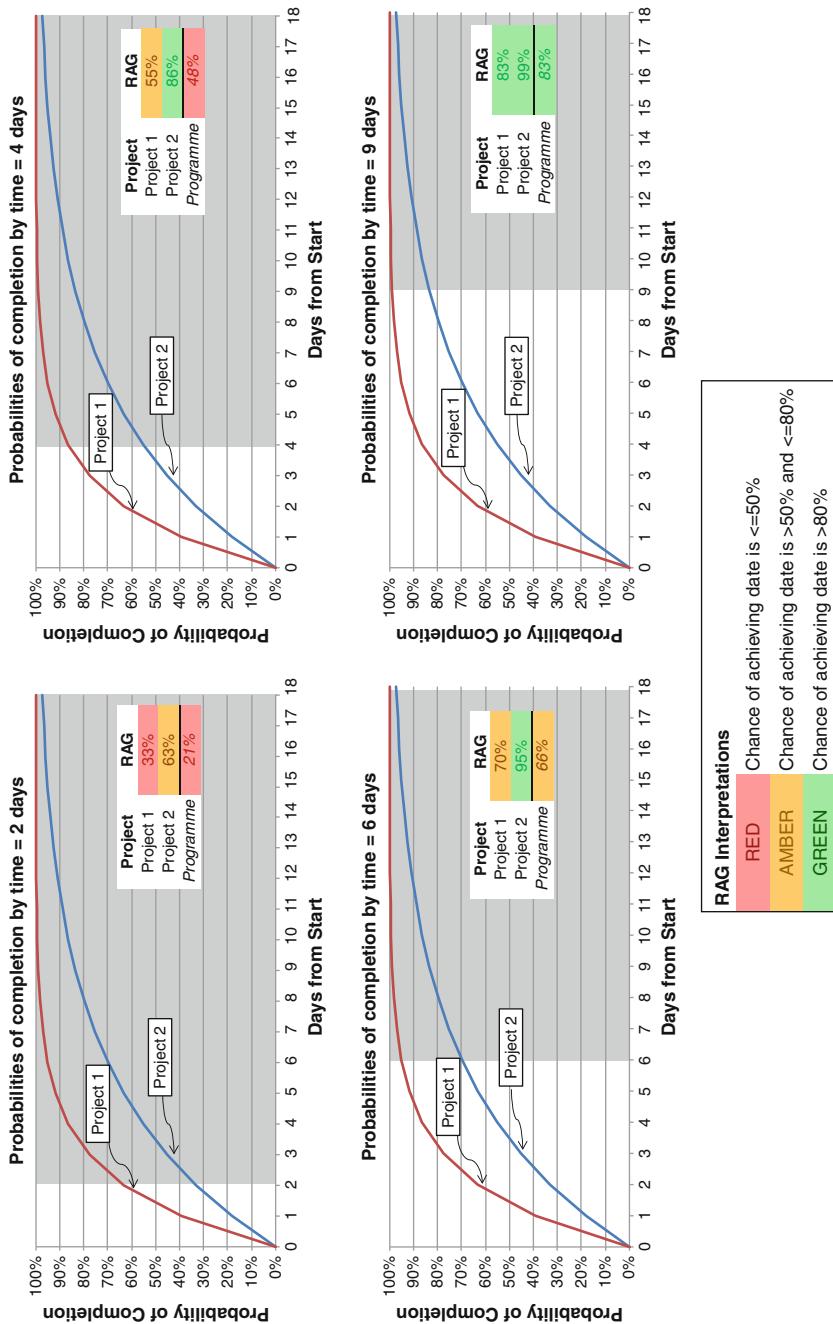


Fig. 17.5  
Project probabilities

The probabilistic method described is not the only way used to combine projects as a combination rule might simply be something like: "Program is RED if any project is RED or more than one Project is AMBER, Program is GREEN if all Projects are GREEN, else Program is AMBER." The rule can be set up as a simple calculation but does not have a ready interpretation in terms of underlying probabilities.

The conclusion is that RAG statuses are subordinate to the calculation of an underlying measure. Given rules for combination of the underlying measures, the combined default RAG status can be set, but there may well be reasons to override the computed RAG such as ROB factors.

Another simple RAG method is to take normalized weighted averages. Here, an RAG represents a banding of measures of our confidence in a certain sense, and the Projects may be "weighted" by their importance. The calculation then takes the sum of the weights times the mean point of the RAG band. For example:

RAG Status bands: GREEN = less than or equal to 1

AMBER = greater than or equal to 1 and less than 2

RED = greater than or equal to 2 and less than 3.

Project A has a weight of 2 and Project B a weight of 10.

If Project A is GREEN and Project B is AMBER, then the Program has RAG status:  
 $(0.5*2 + 1.5*10)/(10 + 2) = 16/12 = 1.33 = \text{AMBER}$

This linear weighting does not allow for the two dimensions of risk measurement – the play-off of chance and harm, or the judgment that associated the outcome distribution of chance and harm to RAGs.

To extend project management assessments to take account of the potential downside, we can introduce a measure of harm, to give what might be termed an *Impact Grid* assignment of RAG to the banded values of chance and harm, as shown in Fig. 17.6:

These assignments of RAG colors to the cells can be set up independently for each Project and for the Program, though there would generally be some level of consistency in what is

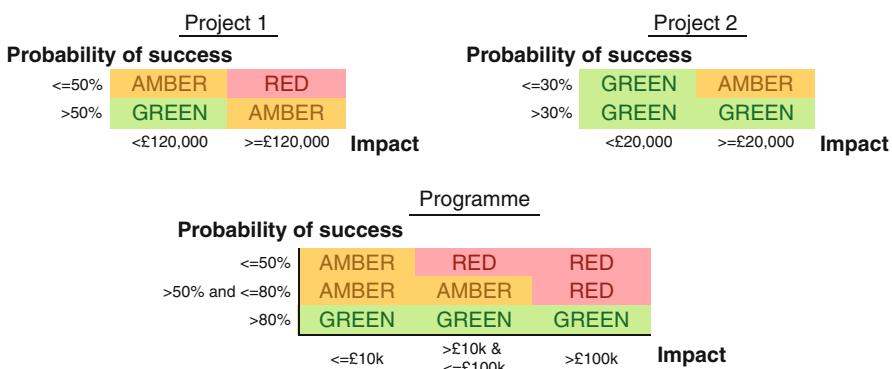


Fig. 17.6

Project risk impact matrices

considered important. However, if Project 2 above had major political implications then it might be set to RED unless the confidence of success were greater than 90%!

The method described works for binary outcome measures like project or task completion. However, the chance versus harm *curves* we have been considering in the theory of risk do not generally simplify to single points representing the risk. The principle still works, though, provided we can assign RAG colors to the chance/harm curves.

➊ *Figure 17.7* shows how this can work. There is an assignment from the risk probability curve using, say, the mean loss and associated chance, to the RAG status based on some quadrant chart of RAGs, against which each Risk can be rated, and then the Overall Risk can be rated in the same way from its composite risk probability curve.

## Solvency II

Solvency II (see, for example, Solvency II 2010) is the EU's regulatory requirement, due to be implemented in December 31, 2012, for the management of risk in the insurance industry notably, but not exclusively, to ensure the capital sufficiency of each insurer. There are many regulator and adviser web sites (e.g., <http://www.lloyds.com/The-Market/Operating-at-Lloyds/Solvency-II>) and books (e.g., Cruz 2009) on the subject, so this section will not recap the whole subject. Instead, we will look at one particular facet, the measurement of risk.

Solvency II adopts that simple and rational measure of risk, VaR, which we came across before, and sets the level of capital required to be 99.5% in a year or allowing a default once in 200 years. This is cautious, for sure; how can it not be sensible? How can it not be in everyone's interest to measure risk in the same way?

As seen in our earlier discussion, however, a VaR measure at low probabilities can be subject to high levels of uncertainty and overdependence on models. The Turner Report and many commentators highlighted model risk from VaR as a major weakness contributing to the banking crisis.

Another critical worry is that risk decisions are not made solely in the light of a once in 200 potential for loss. Otherwise no one would be in business. Instead, business is governed by risk and reward and deciding what to do is a judgment sometimes referred to as *risk appetite*. We are all familiar with this in our daily lives, and it is not that different in business. In Solvency II, however, the measure of risk is not variable, it is fixed. What we then have are two measures of risk which are likely to conflict – the regulatory risk and associated capital to protect society against failure of the company to only 1 year in 200, and the economic risk and associated capital by which a company decides to measure its performance.

➋ *Figure 17.8* shows how this works for the example Classes of Business we looked at earlier, constrained by their VaR for 99.5% being the same (\$30 m) and their mean profit being the same (\$8 m) ignoring cost of capital, and for which the interest charge of the capital is 20% pa (say). For instance, CoB1 might be a less risky line of business such as household fire, and CoB2 might be a riskier line such as US Property Catastrophe.

What we see clearly in the Risk–Reward plot is that if the insurer had a “risk appetite” corresponding to a 25 year return period (and many would have such a horizon) then Property Catastrophe would deliver them a better performance than household fire, but if they chose to be ultra-cautious and adopted a 500-year return period, then they would favor household over catastrophe.

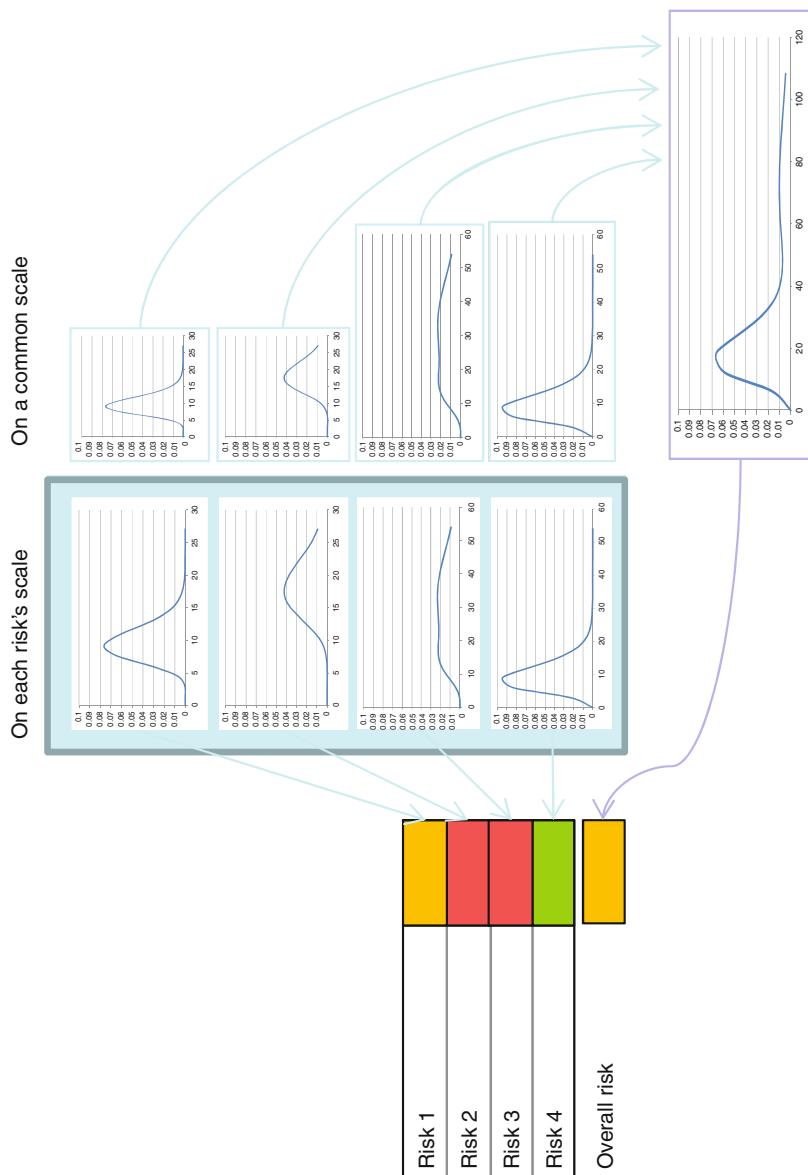
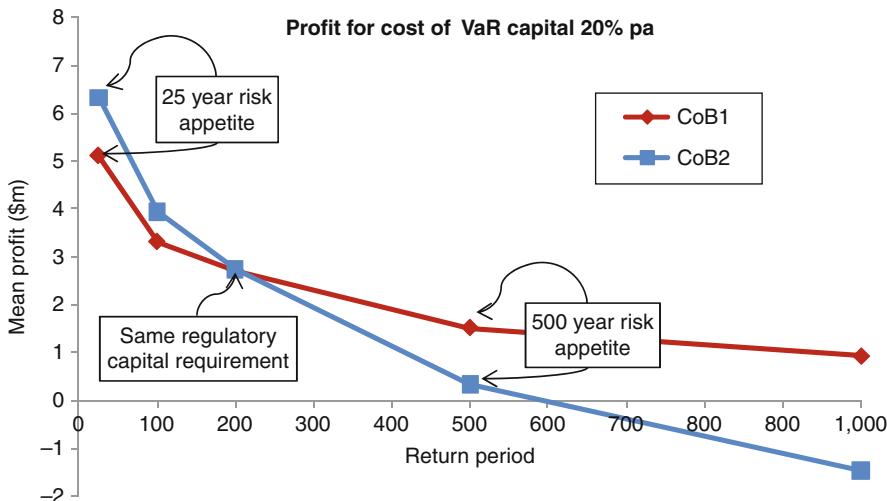


Fig. 17.7  
Combining risk RAGs in general



**Fig. 17.8**  
Risk-reward

The regulatory VaR risk measure is a constraint on the problem and not the risk measure most insurers will be using to manage their business performance.

## Summary

This chapter has set out arguments that risk is a multidimensional concept that cannot in general be reduced to simple measures such as “VaR” used in finance. Risk is proposed to have five dimensions – Time, Harm, Chance, Reward, and Judgment:

- A *Time* frame is a prerequisite of risk assessment sometimes overlooked or treated as implicit when it is of fundamental importance in the ability to anticipate or predict risk.
- *Harm* is downside and can be measured variously – such as mortality, morbidity, or a form of financial loss.
- *Chance* is frequency or probability depending on your point of view. The chapter adopts a “Bayesian” view in that practical risk assessments rest on models of reality so that probability is subjectively contextualized within that model.
- *Reward* is the counterbalance to harm that informs decision-making, sometimes made more complicated when one party’s harm is another’s reward.
- *Judgment*, also termed “Risks Outside the Box” (ROB), allows qualitative assessments that are not covered by the other dimensions and typically represent incompleteness or inadequacy of the model(s) chosen for the other four factors.

Reward and Judgment form part of decision-making about risk rather than risk in itself.

Illustrations of risk assessment, such as project management and Solvency II, are shown to exhibit these five dimensions, with the fifth – Judgment – perhaps the most important of all as it can express concern about the adequacy of any model or models representing the reality of the circumstances under consideration.

This highlights a primary concern in risk assessment, model risk – that is, the risk that we have the wrong model – and finding ways to express this form of uncertainty both qualitatively and, where possible, quantitatively.

## Further Research

---

That as a society we are still very poor at dealing with risk and its importance means this is a hugely popular area, with new developments all the time. Of relevance to the ideas proposed in this chapter are:

1. Developing an evidential factual basis for the many risks that are not readily quantifiable at present due to lack of historical evidence. For example, facts about the volcanic dust and its potential for damage to aircraft engines would have avoided the blind reactions when Eyjafjallajökull erupted. Similarly, in finance, databases on actual operational risks would improve everyone's attempts at quantitative estimates for these traditionally qualitative risks.
2. Shared computer modeling and database capabilities, where models of risk could be run over anonymized databases to compare to benchmarks and to test adequacy of capital and decisions about the business that relate to different risk appetites.
3. Methods for quantifying *model risk* could be developed throughout the insurance and banking industries, as well as in the scientific analysis of complex systems, such as climate change models.
4. Quantitative methods could be developed for combinations of RAG Statuses and use in project and risk management.

## Acknowledgments

---

The “Risk” section was adapted from “What is Risk?” (unpublished) by Rafaela Hillerband and Peter Taylor developed for the Risk Seminar Series at the James Martin School in Hilary Term 2008. I would also like to thank Rafaela for her editorial comments on the manuscript.

My thanks to David Shipley for the Terrorism Risk illustration, to Milan Vukelic and Andrew Baddeley for discussions on the use of reverse stress tests in insurance, to Professor Aurora Plomer for introducing me to Denis Noble’s “Music of Life” and the “Hand Formula,” and Chris Taylor, Milan Vukelic, David Shipley, Henry Ashton, Ben Matharu, and John Thirlwell for kindly commenting on drafts of the paper. Most of all, though, to Ian Nicol for his careful reading, questioning, and suggestions of re-expression for which the reader can be most grateful!

## References

---

- Adams J (1995) Risk. UCL Press, London
- AIR (2003) Secondary uncertainty in AIR models 24 July 2003. Technical Note, Applied Insurance Research, Boston
- Alfred R (2010) June 22, 1783: Icelandic volcano disrupts Europe's economy <http://www.wired.com>thisdayintech/2010/06/0621laki-iceland-volcano-ash-chokes-europe/>
- Atkinson D, Peijnenburg J (2006) Probability without certainty? Foundationalism Lewis Reichenbach debate. *Stud Hist Philos Sci* 37:442–453
- Basel II (2010) [http://www.basel-ii-accord.com/ Basel\\_ii\\_644\\_to\\_682\\_Operational\\_Risk.htm](http://www.basel-ii-accord.com/ Basel_ii_644_to_682_Operational_Risk.htm)
- Beck U (1992) Risk society: towards a new modernity. Sage, London

- Bedford T, Cooke J (2001) Probabilistic risk analysis. Cambridge University Press, see the reference quoted therein – Versteeg M (1977) Estimating common cause failure probabilities in reliability and risk analyses. *J Hazard Mater* 17:215–221
- Beinhocker ED (2006) Complex systems and the origin of wealth. Random House Business, London
- Bernstein PL (1996) The remarkable story of risk. Wiley, New York
- Cooke Y (2010) Yvonne Cooke's article in The Independent on 6th July 2010
- Cruz M (ed) (2009) The Solvency II handbook. Risk Books, London
- Economist (2010) The Economist, May 2010
- Einhorn (2008) Private profits and socialized risk, GARP Risk Review (June/July 2008)
- Ellsberg D (1961) Risk, ambiguity and the savage axioms. *Q J Econ* 75:643–669
- Gould S (1985) The mismeasure of man. Penguin, New York
- Gregoriou GN, Hoppe C, Wehn CS (eds) (2010) The risk modelling evaluation handbook. McGraw-Hill, New York
- Grossman PZ, Cearley RW, Cole DH (2006) Insurance, and the learned hand formula. *Law Probability Risk* 5(1):1–18
- Hubbard D (2009) The failure of risk management: why it's broken and how to fix it. Wiley, New York
- Hume D (1748) Enquiry concerning the principles of human understanding. Oxford University Press, Oxford
- JSB (2010) JSB guidelines for the assessment of general damages in personal injury cases. Oxford University Press, Oxford
- Kaplan S, Garrick BJ (1981) On the quantitative definition of risk. *Risk Anal* 1(1):11–27
- Kendrick M (2008) The great cholesterol con. John Blake, London
- Keynes JM (1937) The general theory of employment. *Q J Econ* 51(1):209–223
- Knight FH (1921) Risk uncertainty and profit. Houghton Mifflin, Boston
- Lanchester J (2010) Whoops!: why everyone owes everyone and no one can pay. Allen Lane, London
- Maudsley H (1867) The physiology and pathology of the mind (BiblioBazaar, LLC 2009)
- Montross F (2010) Model mania, GenRe pamphlet
- Noble D (2008) The music of life: biology beyond genes. Oxford University Press, Oxford
- Oxford Economics (2010) The economic impact of air travel restrictions, Oxford Economics. <http://www.oxforeconomics.com/free/pdfs/oeaviationweb10.pdf>
- Pilkey O, Pilkey-Jones L (2006) Useless arithmetic. Columbia University Press, Columbia
- POST (2009) The dual-use dilemma, Parliamentary Office of Science and Technology, Number 340, London
- RMS (2001) RMS Secondary uncertainty methodology. Risk Management Solutions, California
- RMS (2010) Managing terrorism risk. Risk Management Solutions, California
- Rumsfeld D (2002) DoD news briefing – secretary Rumsfeld and Gen. Myers 12 Feb 2002, <http://www.defense.gov/transcripts/transcript.aspx?transcriptid=2636>
- Salmon F (2009) Recipe for disaster: the formula that killed wall street, Wired Magazine 17.03. [http://www.wired.com/techbiz/it/magazine/17-03/wp\\_quant?currentPage=4](http://www.wired.com/techbiz/it/magazine/17-03/wp_quant?currentPage=4)
- Schrödinger E (1935) The present situation in quantum mechanics. In: Wheeler JA and Zurek WH (1983) Quantum theory and measurement. Princeton University Press, Princeton, p. 137
- Shackle GLS (1961) Decision, order and time in human affairs. Cambridge University Press, Cambridge
- Shipley (2009) Probably wrong – misapplications of probability and statistics to real-life uncertainty, presentations given by Taylor P and Shipley D at Oxford in 2008 and 2009.
- Smith L (2010) Uncertainty, ambiguity and risk in forming climate policy, handling uncertainty in science, royal society (Audio recording <http://royalsociety.org/2010-Handling-uncertainty-in-science/>)
- Society R (1992) Risk: analysis perception and management. Royal Society, London
- Solvency II (2010) <http://www.fsa.gov.uk/Pages/About/What/International/solvency/index.shtml>
- Taleb NN (2006) Fooled by randomness. Penguin, New York
- Taleb NN (2007) The black swan. Random House, New York
- Tett G (2009) Fool's gold: how unrestrained greed corrupted a dream, shattered global markets and unleashed a catastrophe. Little Brown, London
- The Stationery Office (2007) Compensation for injury and death (Ogden tables) (ISBN 9-78-011560125-5). [http://www.gad.gov.uk/services/other%20services/compensation\\_for\\_injury\\_and\\_death.html](http://www.gad.gov.uk/services/other%20services/compensation_for_injury_and_death.html)
- Turner (2009) Turner Report, FSA. [http://www.fsa.gov.uk/pubs/other/turner\\_review.pdf](http://www.fsa.gov.uk/pubs/other/turner_review.pdf)
- Wack P (1985) Scenarios: shooting the rapids. Harvard Business Rev 63(6):139–150
- Wilde GIS (1994) Target risk – dealing with the danger of death, disease and damage in everyday decisions. PDE, Toronto
- Woo G (2004) Understanding terrorism risk, The Risk Report Jan 2004



# **18 Unreliable Probabilities, Paradoxes, and Epistemic Risks**

*Nils-Eric Sahlin*

Lund University, Lund, Sweden

Center for Philosophy of Science, University of Pittsburgh,  
Pittsburgh, PA, USA

<i>Introduction</i> .....	478
<i>Rational Decision-Making</i> .....	479
<i>The Paradoxes of Ellsberg and Allais</i> .....	482
<i>Generalizations, Improvements, and Failures</i> .....	485
<i>Epistemic Risk</i> .....	490
<i>Further Research</i> .....	496

**Abstract:** This paper explores the pros and cons of classical theories of rational decision-making and so-called generalized theories of decision-making. It argues that even if they teach us a great deal about rationality and sound decision-making, and are very useful tools, theories of this kind have serious limitations. In particular they breakdown in the presence of epistemic risk and where there is value uncertainty. It is argued that in these situations we need to take a Socratic approach to risk analysis and risk management.

## Introduction

---

Rational decision-making is the easiest thing in the world. Just follow this simple rule:

(MEU) Choose the alternative with maximal expected utility.

Rational decisions are determined by a combination of our beliefs and desires – our beliefs determining the probabilities of the outcomes, our desires determining the utilities of the possible outcomes.

The classical view assumes that the rational decision maker can choose between a set of alternatives. The alternatives lead to different outcomes depending on the state of affairs each alternative brings into being. An ideal decision situation can therefore be described in terms of a simple decision matrix as shown in [Table 18.1](#).

If  $P(\cdot)$  is the probability function defined over states, the states being uncertain and outside the decision maker's control, and if  $u(\cdot)$  is the utility function defined over outcomes, representing the decision maker's preferences, the expected utility of alternative  $a_i$  is  $P(s_1)u(o_{i1}) + P(s_2)u(o_{i2}) + \dots + P(s_m)u(o_{im})$ . This is the value to maximize.

In this paper the classical theories of rational decision will be outlined and some of their shortcomings will be discussed. It will be argued that these theories ask too much of us – e.g., that we have perfect information and knowledge, that there is no epistemic uncertainty whatsoever, and that we have determinate preferences. Taking a couple of well-known problems, or paradoxes, as a point of departure, alternatives to the classical theories will be explored. Generalized theories, in their efforts to model epistemic uncertainty, give up one or more of the classical axioms. Examples will show that this leads to various types of more or less serious problem. Examining some contemporary risk debates (e.g., nanotechnology, stem cell research, synthetic biology), we clearly see that the evaluation of epistemic uncertainty is as problematic as assessment of the disutility of potential negative outcomes. The epistemic uncertainty creates epistemic risks. However, if our formal theories of rational decision-making cannot guide us, how do we then make decisions? A Socratic approach to decision-making and risk analysis is advocated.

**Table 18.1**  
A decision matrix

Alternatives	States			
	$s_1$	$s_2$		$s_m$
$a_1$	$o_{11}$	$o_{12}$		$o_{1m}$
$a_2$	$o_{21}$	$o_{22}$		$o_{2m}$
...	...	...		...
$a_n$	$o_{n1}$	$o_{n2}$		$o_{nm}$

## Rational Decision-Making

Several theories can be called classical theories of rational decision-making; all ask us to maximize expected utility in one way or another. The best known is probably L. J. Savage's theory, presented in *The Foundations of Statistics* (1954); the first complete theory is probably F. P. Ramsey's, presented in the paper "Truth and probability" (1929). Other examples of classical theory are Debreu (1959), Anscombe and Aumann (1963), and, more controversially, Jeffrey (1965). All outlined and discussed in Fishburn (1981). Here I want to briefly contour Ramsey's theory, sketching his aims and intentions and assessing what he achieves.

Ramsey showed that, under ideal conditions, people's beliefs and desires can be measured with a betting method, and that, given some intuitive principles of rational behavior, a measure of our "degrees of belief" will satisfy the laws of probability: that is, he gave us the modern theory of subjective probability. He was the first to state the Dutch book theorem, and he laid the foundations of modern utility theory and decision theory. In addition, he had a proof of the value of collecting evidence years before it became known through the independent work of Savage and I. J. Good (1967) (see also Sahlin's preamble to Ramsey 1990); he took higher order probabilities seriously; and, in a derivation of the "rule of succession," he introduced the notion of "exchangeability" (under a different name, see M. C. Galavotti's (1991) edition of Ramsey's notes). Ramsey's decision/probability theory is just about as complete as any such theory could be (see Sahlin 1990).

The aim of "Truth and Probability" is to analyze the connection between the subjective degree of belief we have in a proposition and the (subjective) probability it can be assigned, and to find a behavioral way of measuring degrees of belief. Ramsey shows: first, that we can measure the degree of belief a subject has in a given proposition; and, second, that if the subject is rational, his or her degrees of belief will have a measure satisfying the axioms of probability theory – a "subjective" or "personal" probability. In other words, Ramsey shows that, given his method of measuring the strength of "partial beliefs," the degrees of belief displayed by an ideally rational decision maker will be ruled by the laws of probability.

So how does Ramsey propose to measure people's beliefs and desires with a traditional betting method? We can measure a subject's belief simply by proposing a bet. We must "see what are the lowest odds which he will accept" (Ramsey 1926/1990, p. 68. All page references to the 1990, Mellor edition, *Philosophical Papers*, of Ramsey's works). The strategy is to offer the decision maker a bet on the truth value of the proposition  $p$  believed. Ramsey took this method to be "fundamentally sound," but argued that it suffers from "being insufficiently general, and from being necessarily inexact . . . partly because of the diminishing marginal utility of money, partly because the person may have a special eagerness or reluctance to bet. . ." (p. 68).

A bet is of the form:  $x$  if  $p$  is true,  $y$  if  $p$  is not true, where  $x > y$ . The "traditional method" tells us that the decision maker's degree of belief in  $p$  is  $(f - y)/(x - y)$ , where  $f$  is the greatest amount the decision maker is willing to pay for the bet. Notice that the smallest amount of money the decision maker would be prepared to pay for the bet coincides with the smallest amount for which the decision maker would sell it. If the bet is in money and the marginal utility for money is decreasing, it is obvious that using monetary outcomes will fail to give correct measures for (say) bets involving substantial sums of money. As Ramsey says the betting method, though "sound," is neither completely "general" nor very "exact."

Ramsey equates a degree of belief of  $\frac{1}{2}$  in an ethically neutral proposition  $p$  with indifference to two options:  $a$  if  $p$  is true,  $b$  if  $p$  is not true; and  $b$  if  $p$  is true,  $a$  if  $p$  is not true ( $a, b, c, \dots$ ,

denoting outcomes). “This comes roughly to defining belief of degree  $\frac{1}{2}$  as such a degree of belief as leads to indifference between betting one way and betting the other for the same stakes”(p. 74). An ethically neutral proposition of degree  $\frac{1}{2}$  comes close to something like a platonic idea of a fair coin.

This gives Ramsey a way of measuring value differences. Thus the notion that the value difference between  $a$  and  $b$  is equal to the difference between  $c$  and  $d$  simply means that if  $p_{\frac{1}{2}}$  is an ethically neutral proposition believed to degree  $\frac{1}{2}$ , the options ( $a$  if  $p_{\frac{1}{2}}$  is true, and  $d$  if  $p_{\frac{1}{2}}$  is not true) and ( $b$  if  $p_{\frac{1}{2}}$  is true, and  $c$  if  $p_{\frac{1}{2}}$  is not true) are equally preferable.

Ramsey proves a representation theorem saying that a subject’s preferences can be represented by a utility function determined up to a positive linear transformation. It is the binary preferences that are represented, and the very goal of the theorem is to isolate the conditions under which such preferences can be seen as maximizing expected utility. The representation guarantees the existence of a probability function and an unconditional utility function such that the expected utility defined from this probability and utility represents the decision maker’s preferences. To prove this theorem eight axioms are introduced. The axioms, following Suppes (1956), can be divided into three groups: behavioral, ontological, and structural.

A behavioral axiom is a rule that a rational person is supposed to satisfy when making a decision. One of Ramsey’s axioms states that the subject’s value differences are transitive: If the difference in value between  $a$  and  $b$  is equal to the difference between  $c$  and  $d$ , and the difference between  $c$  and  $d$  is equal to that between  $e$  and  $f$ , then the difference between  $a$  and  $b$  is equal to that between  $e$  and  $f$ . This is a typical behavioral axiom.

The ontological and structural axioms tell us what there is and give us the mathematical muscles necessary to prove the representation theorem. Ramsey’s first axiom, for example, states that “[t]here is an ethically neutral proposition  $p$  believed to degree  $\frac{1}{2}$ ” (Ramsey 1926/1990, p. 74). As it turns out, this ontological axiom is far more important than one might have expected. Ramsey’s two final axioms are structural: one is an axiom of continuity, the other an Archimedean axiom.

Ramsey’s utility theory is closely related to the theory developed by von Neumann and Morgenstern in *Theory of Games and Economic Behavior* (1944/1953) about two decades later. von Neumann and Morgenstern, however, use objective or stated probabilities, prizes, and lotteries to derive the utilities. Ramsey avoids these assumptions. As a result he need not postulate that subjects understand the information contained in a stated probability, nor need he assume that the information is well calibrated (i.e., that the subjective probabilities mirror the stated objective probabilities). In a betting method with value differences a number of problems are addressed. The worry that the traditional method, involving monetary outcomes, is insufficiently general and necessarily inexact no longer arises, and concerns about the diminishing marginal utility of money and risk-aversion (risk-proneness) fall away.

It is then possible to define the degree of belief in  $p$  “by the odds at which the subject would bet on  $p$ , the bet being conducted in terms of differences of value as defined” (Ramsey 1926/1990, p. 76). If the subject is indifferent between  $a$  with certainty, and  $b$  if  $p$  is true ( $p$  not necessarily being an ethically neutral proposition, although  $p$ ’s truth cannot change the relative values of the outcomes) and  $c$  if  $p$  is not true, the subject’s degree of belief in the proposition is defined as the difference in value between  $a$  and  $c$  divided by the difference in value between  $b$  and  $c$ .

This can be expressed formally as follows:

$$P(p) = (u(a) - u(c)) / (u(b) - u(c)),$$

where “ $P(.)$ ” denotes the subject’s degree of belief function and “ $u(..)$ ” denotes the subject’s utility function. Ramsey also shows how the degree of belief in a proposition, given the truth of another proposition, can be defined along the same lines, using a slightly more complicated pair of bets.

Ramsey then proves that the obtained measure of degree of belief is a probability measure – it obeys the axioms of probability theory. The probability of any proposition is greater than or equal to 0; the probability of a proposition plus the probability of its negation equals 1; and, if two propositions are incompatible, the probability of the disjunction equals the sum of the probability of the disjuncts. Furthermore, Ramsey proves the Dutch book theorem: “[h]aving degrees of belief obeying the laws of probability implies a further measure of consistency, namely such a consistency between the odds acceptable on different propositions as shall prevent a book being made against you” (Ramsey 1926/1990, p. 79). Having degrees of belief obeying the axioms of probability, having a coherent set of beliefs, is simply a logically necessary and sufficient condition of avoiding a Dutch book. It should be noted that a subject can have more or less any degree of belief whatsoever in a proposition provided the set of beliefs to which it belongs is coherent (consistent). It is essentially this feature of Ramsey’s theory that makes the theory subjectivist.

As mentioned above, Ramsey’s decision/probability theory is as close to being complete as any such theory can be. That is to say, it is complete in the sense that it deals with, and provides answers to, the fundamental questions we have – for example, the question answered above in the prolegomenon: How do we make a rational decision?

In the present context it is the rationality rules that are of particular interest. Basically there are two of them: ordering assumptions and independence assumptions (see Seidenfeld 1988).

Transitivity, which is an ordering assumption, is one kind of a rationality rule. Rational Man’s preferences are assumed to be transitive. This is something we, too, want our preferences to be, since if they are not, we risk becoming money pumps.

The independence axioms of decision theory and game theory – e.g., Savage’s Sure-Thing Principle – are another classical type of rationality axiom. Savage’s principle tells us that if an alternative  $A$  is judged to be as good as another  $B$  in all possible states and better than  $B$  in at least one, then a rational decision maker will prefer  $A$  to  $B$ . Savage (1954/1972) illustrates the principle with the following case:

- ▶ A businessman contemplates buying a certain piece of property. He considers the outcome of the next presidential election relevant to the attractiveness of the purchase. So, to clarify the matter for himself, he asks whether he would buy if he knew that the Republican candidate were going to win, and decides that he would do so. Similarly, he considers whether he would buy if he knew that the Democratic candidate was going to win, and again finds that he would do so. Seeing that he would buy in either event, he decides that he should buy, even though he does not know which event obtains, or will obtain, . . . [E]xcept possibly for the assumption of simple ordering, I know of no other extralogical principle governing decisions that finds such ready acceptance (p. 21).

Similarly, moving to game theory, and assuming stated probabilities, the so-called strong independence axiom tells us that for all outcomes  $A$ ,  $B$ ,  $C$ , and probability  $p > 0$ ,  $A$  is better than  $B$  if and only if a prospect  $A$  with probability  $p$  and  $C$  with probability  $1 - p$  is better than a prospect  $B$  with probability  $p$  and  $C$  with probability  $1 - p$ .

Savage formulated the Sure-Thing Principle in 1954 in *The Foundations of Statistics*. By his own account, his view of probability derives mainly from the work of de Finetti (1937, 1974) and is inspired by Ramsey's theory. Savage's theory is in many respects similar to Ramsey's, but instead of giving us a narrative of an ideal he presents a normative theory: that is, Savage's theory tells us how we *ought* to choose. Ramsey's theory, by contrast, is descriptive. If we assume, however, that the decisions of the ideal decision maker, Rational Man, are the best that can be made, and that we human beings want to and maybe should follow in his, her, or its footsteps, a normative interpretation of Ramsey is not far away.

The Sure-Thing Principle and the ordering axioms are accepted by many (perhaps most) normative decision theorists (see Fishburn 1981 for a review of normative theories). Though they sometimes appear counterintuitive (we will discuss a few paradigm cases of this below), they are considered cornerstones of rationality that we ought to follow.

But be aware that even if the classical theories have much in common, especially in their view of rationality, they are not identical. Thus the ontological and structural axioms are different. Ramsey is working with propositions and words (outcomes or consequences), Savage with events and consequences. Savage defines acts as a function from states to consequences, which, when set against Ramsey's theory, is an innovation. And also small differences can matter.

Schervish et al. (1990) have shown that classical theories such as those developed by Savage, von Neumann and Morgenstern, Anscombe and Aumann, and de Finetti all have a problem with state-dependent utilities. Theories using (horse) lotteries and prizes to derive probabilities cannot guarantee the existence of unique probabilities. The problem is that the utility of a prize is the utility of that prize given that a particular state of nature obtains. And even "constant" prizes might have different values in different states of nature – with the implication that the subject's preferences can be represented by far too many utility functions. As a consequence there is no unique subjective probability distribution over states of nature. Ramsey saw that his method of using preferences among bets to quantify value differences required the states defining the bets to be value-neutral. For that, he proposed "ethically neutral propositions." In Ramsey's theory the outcomes have state-dependent utilities, which can be measured through bets involving an ethically neutral proposition. The question, however, is whether the concept of an ethically neutral proposition (Axiom 1: "There is an ethically neutral proposition believed to degree  $\frac{1}{2}$ ") can be understood and pressed into service without our making use of lotteries. If not, Ramsey faces the same problems as did the descendants of his theory; but if it can be done, Ramsey's theory has a clear advantage over the other classical theories.

## The Paradoxes of Ellsberg and Allais

---

In a classic paper Daniel Ellsberg (1961) asks us to consider a decision problem. Imagine an urn containing 30 red balls and 60 that are either black or yellow, in unknown proportions. A ball is to be drawn at random from the urn. You are asked to make two choices. The first choice is between alternatives  $e_1$  and  $e_2$ . If you choose  $e_1$  you will receive \$100 if a red ball is drawn and nothing if a black or a yellow ball is drawn. If you choose  $e_2$  you will receive \$100 if a black ball is drawn, nothing if a red or yellow ball is drawn. The second choice is between alternatives  $e_3$  and  $e_4$ . If you choose  $e_3$  you will receive \$100 if red ball is drawn, \$100 if a yellow ball is drawn, and nothing if a black ball is drawn. If you choose  $e_4$  you will receive \$100 if a black ball is drawn, \$100 if a yellow ball is drawn, and nothing if a red ball is drawn.

The problem can be represented by a decision matrix as shown in [Table 18.2](#).

Studies have shown that a vast majority of subjects prefer  $e_1$  to  $e_2$ , and  $e_4$  to  $e_3$  (see Kahneman and Tversky 1979; Sahlin 1991). This pair of preferences violates the Sure-Thing Principle. Note that  $e_1$  is the same alternative as  $e_3$ , except for the sure-thing: \$100 if yellow. And  $e_2$  is identical to  $e_4$ , except for the sure-thing: \$100 if yellow.

It can also be argued that preferring  $e_1$  to  $e_2$  (or for that matter  $e_4$  to  $e_3$ ) is, in isolation, irrational. Ramsey might, and Savage definitely would, argue that the two alternatives have the same expected value, i.e., \$33. The uncertainty when it comes to the proportion of black and yellow balls is of no significance. The exact proportion is unknown and it is thus reasonable to assume that all possible proportions have the same probability (applying a principle of indifference). The proportion of black and yellow balls is a random variable, with respect to a known probability, and the *expected* proportion of balls is 30-30-30. This is, to use Savage's phrase, the "composite probability" that ought to be used in rational decision – which implies, in this case, that  $e_1$  is as good as  $e_2$ .

It is important to note that Ellsberg's paradox teaches us two things. First, the reliability or unreliability, and the determinacy or indeterminacy, of probabilities influences our actions. Second, violating the independence assumption expressed in the Sure-Thing Principle sometimes seems perfectly rational – does it not? There must be theories that prescribe and describe the Ellsberg types of choice. Or, is the only option for those of us who choose  $e_1$  and  $e_4$  to confess irrationality and resolve to work hard to become a better person, i.e., adapt our decisions to the norm?

Maurice Allais (1953) attacks the classical theory of rational decision-making with a nice example known, in popular parlance, as Allais's paradox.

You have two choices. The first choice is between alternatives  $a_1$  and  $a_2$ , the second between  $a_3$  and  $a_4$ . If you choose  $a_1$  you get nothing (\$0) with probability 0.01, a nice round sum of \$2,500,000 with probability 0.10, or \$500,000 with probability 0.89. If you choose  $a_2$  you get \$500,000 with certainty.

The alternatives  $a_3$  and  $a_4$  have the following outcomes. If you choose  $a_3$  you get nothing with probability 0.90 and \$2,500,000 with probability 0.10. If you choose  $a_4$  you get \$500,000 with probability 0.11 or nothing with probability 0.89.

The paradox can be represented by a decision matrix as shown in [Table 18.3](#).

Most people prefer  $a_2$  to  $a_1$ , but  $a_3$  to  $a_4$ . As in Ellsberg's paradox it is the sure-thing, in this case the \$500,000 with probability 0.89, that causes the problem. Bar the sure-thing,  $a_1$  would have given the same outcomes as  $a_3$  and  $a_2$  would have given the same outcomes as  $a_4$ . If  $a_3$  is, so to speak, better than  $a_4$  without the sure-thing, why is it not so with the sure-thing?

This type of "irrational" choice is dubbed "the certainty effect," by Kahneman and Tversky (1979). It appears to be the best known violation of expected utility theory. In controlled

**Table 18.2**  
**Ellsberg's paradox**

	Red	Black	Yellow
$e_1$	\$100	Nothing	Nothing
$e_2$	Nothing	\$100	Nothing
$e_3$	\$100	Nothing	\$100
$e_4$	Nothing	\$100	\$100

■ **Table 18.3**  
**Allais's paradox**

	0.01	0.10	0.89
$a_1$	\$0	\$2,500,000	\$500,000
$a_2$	\$500,000	\$500,000	\$500,000
$a_3$	\$0	\$2,500,000	\$0
$a_4$	\$500,000	\$500,000	\$0

empirical studies Kahneman and Tversky have investigated the following version of Allais's paradox. Their subjects were asked to choose between:

$A_1$ :	2,500	With probability 0.33
	2,400	With probability 0.66
	0	With probability 0.01

and

$A_2$ :	2,400	With certainty
---------	-------	----------------

And to choose between:

$A_3$ :	2,500	With probability 0.33
	0	With probability 0.67

and

$A_4$ :	2,400	With probability 0.34
	0	With probability 0.66

Their empirical finding was that 82% of the subjects preferred  $A_2$  to  $A_1$  and that 83% preferred  $A_3$  to  $A_4$ . Together these choices violate expected utility theory. That is, the pair of preferences implies a pair of incompatible utility inequalities:  $0.34u(2,400) > 0.33u(2,500)$ , by the first preference, but, at the same time, the reverse inequality holds by the second preference. In Allais's paradox, given the preference of  $a_2$  over  $a_1$ , we obtain  $0.11u(\$500,000) > 0.10u(\$2,500,000)$ , an inequality reversed by the preference of  $a_3$  over  $a_4$ . In both cases we have to assume that  $u(\text{nothing}) = 0$ , and that subjective probability assessments reflect stated probabilities.

Note that to secure a violation of the independence assumption expressed in the Sure-Thing Principle we must assume that we have the same utility function in both choice situations – that is, that one and the same utility function is in play when we choose between  $a_1$  and  $a_2$  (or  $A_1$  and  $A_2$ ) and when we choose between  $a_3$  and  $a_4$  (or  $A_3$  and  $A_4$ ). The violation lapses if we assume, or can show, that we are employing distinct utility functions in the two decision situations (Sahlin 1991). Introspection tells me that I, for one, do not have the same utility function in the two

choice situations. In the first situation I know that I can with certainty collect \$500,000 (or 2,400). I therefore evaluate the \$2,500,000 (or 2,500) outcome in the light of this possible prosperity, rather than in the light of my present wealth. The \$500,000 (or 2,400) changes my levels of aspiration and thus my utility assessments. In the second choice situation there is no choice which, with certainty, will give a small fortune in dollars. I, and you, have to evaluate the possible outcomes with respect to our present wealth. This simple observation makes the empirical findings on Allais's paradox less robust and therefore less interesting perhaps than similar results obtained in connection with the Ellsberg paradox (Sahlin (1988)).

Ellsberg's paradox and Allais's paradox have been very influential. Two relatively simple thought experiments have made decision theoreticians lose faith in the commandments of the classical theory. They have triggered numerous of empirical studies and influenced many descriptive theories (including Prospect Theory, see Kahneman and Tversky 1979), and they have made us question our rationality – we are irrational and only ideal decision makers are rational (see Bränmark and Sahlin 2010; Sahlin and Bränmark 2011). They have also inspired the development of new, generalized “normative” decision theories. And this is the topic of the following section.

## Generalizations, Improvements, and Failures

---

Ramsey's, Savage's, and other classical theories seek to represent a subject's preferences with a utility function determined up to an affine transformation, and a subject's state of belief by a unique probability measure, to any degree of precision. Ellsberg's paradox – and to some extent Allais's paradox – shows that this aim leads to assumptions that are far too strong. The assumptions made by the classical theories are also put in question by numerous of empirical findings (if normative theory should bother about the trifles of everyday life, see Sahlin 1988, 1991) and by philosophical arguments and questions. What is a decision maker? Does a decision maker have, in addition to full beliefs and degrees of belief, higher order beliefs? Should the normative theory be designed for an ideal, unrealistic agent or a more plethoric decision maker? What is our aim: large inapplicable norms, or far from faultless prescriptive theories, with known blemishes, the good supervisor (see Sahlin and Vareman 2008).

It is not too difficult to calculate the probability that a white ball will be drawn from an urn containing 99 white balls and 1 black ball. It is far more difficult to estimate, with reasonable precision, the probability that my friend in Canberra is at this very moment drinking tea. And it is almost impossible, in the light of what we know today, to say with the degree of precision required by the classical theories what the probability is that, if they are transplanted, iPS-cells (induced pluripotent stem cells) will behave the same way as ePS-cells (embryonic stem cells). There simply are too many things we know too little about, like undiscovered epigenetic factors (Sahlin et al. 2010).

It is well known that in choice situations degrees of indeterminacy or unreliability may influence decisions. We may, for example, think that the probability is the same (say 0.99) in all three of the examples above, but still, if we have a choice, we prefer to bet on a fair gamble if the outcome is determined by the drawing from the urn, but not if it is determined by what is going on in Australia or the advancements of stem cell research. We have a feeling that we know more about the urn than we do about our friends' habits and the very latest findings in bioscience. The stability of our knowledge influences our decision-making. In other words, there are situations in which there are important

differences of degree in our knowledge, or understanding, of the various factors underlying our decisions – a difference in ignorance that cannot be mirrored by a unique probability measure, or captured in the classical theories, or for that matter seeds of those theories (detectable in Jeffrey 1965 and Bolker 1967).

The probabilities given by Ramsey's theory (and similar theories like Savage's, Anscombe and Aumann's, and Jeffrey's) are the result of the decision maker's inability to express fully a strength of preference. Ramsey's theory does not take unreliability into account. He discusses higher order probabilities, but not as an integrated part of the 1926 theory (Sahlin 1993). The necessary clarity of perception of uncertainty (caused, in part, by the quantity and quality of information with which the decision maker is equipped) is simply not introduced.

It appears, then, that it is necessary to create theories that allow indeterminate, or unreliable, probabilities and utilities so that those theories can be applied to real-life decision problems. But, as it turns out, this improvement is not an uncomplicated matter. A comparison of two theories, two attempts to solve the problem, will reveal some of the difficulties with this type of generalized decision theory.

In a number of important works Isaac Levi (1980, 1986, 1988) has developed a decision theory permitting the decision maker indeterminate probabilities and imprecise preferences. Levi assumes that the decision maker  $X$ 's knowledge and information at time  $t$  about states of nature is contained in a convex set  $B_{X,t}$  of probability distributions.  $B_{X,t}$  is the set of *permissible* distributions. Levi's approach provides an impressively adequate and complete representation, allowing the quality and quantity of the decision maker's information to be more fully accounted for.

Levi also generalizes the classical theory by introducing a set of “permissible” utility functions (denoted  $G$ ). The classical theory assumes that there is but one utility function representing the decision maker's preferences (not counting affine transformations, which anyway are not countable). A set of utility functions tells us far more about the uncertainty and robustness of the decision maker's preferences than a single function does.

As long as the decision maker's beliefs and values can be represented by a unique probability distribution and a single utility function it is possible to maximize expected utility. But with sets of measures the situation is different. We cannot maximize over intervals or sets. Generalized decision theories therefore have to offer new decision rules. Levi suggests we use a three-step procedure, a lexicographically ordered set of carefully chosen rules, of the following sort.

First, a decision alternative is said to be *E-admissible* if and only if there is some probability distribution in the set  $B_{X,t}$  and some utility function in the set of utility functions  $G$  such that the expected utility of the action alternative relative to the two distributions is maximal among all the available decision alternatives. It is then stated that a decision alternative must be E-admissible in order to be choice-worthy.

An alternative is *P-admissible* if it is E-admissible and it is “best” with respect to E-admissible option preservation among all E-admissible options. This condition has to do with the possibility of deferring decision. “[T]he injunction to keep one's options open is a criterion of choice that is based not on appraisals of expected utility but on the ‘option-preserving’ features of options” (Levi 1980, pp. 156–163).

Finally, a P-admissible alternative is *security-optimal* if and only if the minimum utility value assigned to some possible outcome of the decision alternative is at least as great as the minimal utility value assigned to any other P-admissible alternative. And a decision alternative is S-admissible if it is E-admissible, P-admissible, and security-optimal (relative to some of the

decision maker's utility functions). It is the S-admissible alternatives that are admissible for final choice. If there is more than one admissible alternative, a dice, or similarly random device, has to be introduced.

After Ramsey's work, Levi's theory is an indisputable breakthrough. One of the theory's great innovations is the way it represents the decision maker's beliefs and desires. Another improvement is the carefully developed set of decision rules.

A theory with ambitions similar to Levi's is presented by Gärdenfors and Sahlin (Gärdenfors and Sahlin 1982, 1983; Sahlin 1983, 1985, 1993). The theory assumes that the decision maker's knowledge and beliefs concerning states of nature can be represented by a set  $P$  of probability distributions, which is the set of epistemically possible distributions. This set consists of those measures which do not contradict the knowledge the decision maker has.  $P$  is restricted by way of a second-order measure of reliability. A second-order probability measure,  $\rho$  defined over the set of "first-order" probability measures, allowing for a representation not only of the decision maker's first-order beliefs but also of his or her higher order beliefs. As a basis for action, a decision maker should use those and only those measures of  $P$  that are epistemically reliable, which implies that a subset  $P/\rho_0$  of  $P$  is to be used when making a decision, with  $\rho_0$  being a level of aspiration with respect to epistemic uncertainty. This restricted set can be compared with Levi's set  $B_{X,t}$ . Note, for example, that unlike Levi's set,  $P/\rho_0$  does not have to be convex.

Gärdenfors and Sahlin's theory is a two-step rule for making decisions. First, the expected utility of each choice alternative, and each probability distribution in  $P/\rho_0$ , is computed, and the minimal expected utility of each alternative is determined. (Gärdenfors and Sahlin do not introduce a set  $G$  of utility functions, although to do so would be straightforward.) Second, the choice alternative with the largest minimal expected utility is selected. So, instead of the classical goal of maximizing expected utility, it is suggested that we should *maximize the minimal expected utility* (MmEU).

Gärdenfors and Sahlin's theory "solves" the Ellsberg paradox straight out – or rather, the theory gives a recommendation that accords with Ellsberg's findings. In Ellsberg's example the set of epistemically possible distributions is  $(1/3, x, 2/3 - x)$ ,  $0 \leq x \leq 2/3$ . It is a reasonable assumption that  $P = P/\rho_0$ . The minimal expected utilities for the four alternatives,  $e_1$ ,  $e_2$ ,  $e_3$ , and  $e_4$ , can now be calculated. If  $1/3u(\$100) > u(\$0)$ , then MmEU tells us that  $e_1$  should be preferred  $e_2$ , and  $e_4$  preferred to  $e_3$ . Ellsberg's subjects – and the subjects in, for example, Goldsmith and Sahlin's (1983) experiments – remind us that decision makers' do not just perceive epistemic uncertainty, but are capable of consistently using this uncertainty when making decisions. The alternatives  $e_2$  and  $e_3$  involve greater "epistemic risk," or greater epistemic uncertainty, than  $e_1$  and  $e_4$ , and are consequently (in the present outcome matrix) avoided.

But this solution comes at a cost. Seidenfeld (1988) has constructed a simple but ingenious decision problem. Imagine that Seidenfeld has a stone in one of his pockets. Your task is to predict whether it is in his left or right pocket. If you say Left and the stone is in the left pocket he gives you \$100, and if you say left and it is in the right pocket he gives you \$10. If you say Right the outcomes are reversed; you get \$100 if the stone is in the right pocket and \$10 if it is in the left pocket. Your choices, Left and Right, therefore have the same minimal expected value of \$10. MmEU tells you to be indifferent between the two options.

Now construct a 50:50 gamble with outcomes Left or Right – that is, a 50:50 gamble with outcomes \$10 or \$100. It can be assumed that this gamble has an expected value of \$27.50 (with dollars reflecting utilities).

In a clever move, Seidenfeld now constructs two further gambles. Gamble 1 is this: if a fair coin lands on heads you can either choose Left or \$9 in cash, if it lands tails you can choose either Right or \$9. Gamble 1 has the same value as the 50:50 gamble between Left and Right, i.e., \$27.50. Gamble 2 is this: if a fair coin lands heads you can either choose Left or \$11, if it lands tails you have a choice between Right and \$11. Left and Right have the same expected value, \$10, so you will in both choice situations take the “sure-thing,” the 11 dollars. This means that Gamble 1 is worth \$27.50 and Gamble 2 is worth \$11. So far so good. But note, says Seidenfeld, if the coin lands on heads Gamble 1 gives \$10 but Gamble 2 one dollar more. And so it does if the coin lands on tails. So is not Gamble 2 the obvious choice?

What Seidenfeld’s example shows is that theories like Gärdenfors and Sahlin’s can make counterintuitive recommendations in consecutive choice situations. This is a problem that the theory shares with all normative, prescriptive, and descriptive theories (e.g., Prospect Theory) that sell the independence postulate to obtain explanatory power or rational resilience. Levi’s theory, with its lexicographically ordered set of decision rules, with E-admissibility as the cornerstone, avoids the trap.

But Levi’s theory is not without blemishes. For instance, it does not give us a direct solution to the Ellsberg paradox. Note that  $e_1$  and  $e_2$  are both E-admissible; they are both S-admissible, too. The same holds for  $e_3$  and  $e_4$ . In other words,  $e_1$  is as good a choice as  $e_2$ : both are admissible. And  $e_3$  is as good a choice as  $e_4$ : both are admissible. Levi (1986) explains how this (perhaps unwanted) consequence can be avoided.

The problem with Ellsberg-type choices is the unreliable probability distribution. We simply do not know the proportion of black and yellow balls. We do know, however, that the actual distribution is one of 61 possible permutations. Let  $h_i$  be the hypothesis that  $i$  balls are black. The probability of drawing a black ball given that  $h_i$  is true is  $i/90$ . For each alternative and each distribution the expected value conditional on  $h_i$  can be determined. The two original decision matrixes can now be redrawn. Each alternative,  $e_1$ ,  $e_2$ ,  $e_3$ , and  $e_4$ , now has 61 possible outcomes, not 3. The outcomes are expected values, not fixed monetary prizes.

Levi points out that this transformation of the original problem changes nothing when it comes to E-admissibility. All four alternatives are E-admissible. However, the security values are different.  $e_1$ ’s security level is  $100/3$ ,  $e_2$ ’s security level is 0, and for  $e_3$  and  $e_4$  the levels are  $100/3$  and  $200/3$ , respectively. In the pursuit of security-optimal choices,  $e_1$  should be preferred to  $e_2$ , and  $e_4$  to  $e_3$ . This is an artful transformation of the two original decision matrixes that neatly harmonizes the recommendations given by Levi’s theory with Ellsberg’s findings.

This can be seen as trickery. Levi’s theory states that an alternative is security-optimal if and only if the minimum utility value assigned to some possible *outcome* of the decision alternative is at least as great as the minimal utility value assigned to any other admissible alternative. What is meant by an outcome here? Is an expected value an outcome? Utilities defined over what? We know that it is far more difficult to find counterexamples to preference principles in a probability free world than it is when probabilities are allowed to, as it were, infect values. And if state-dependent utilities are a problem for the classical theory, with the type of outcomes it deals in, Levi’s move introduces new levels of dependence. Another thing one might ask is whether every one of the 61 possible distributions has the same probability. Is it more likely that there are 30 black and 30 yellow balls in the urn than it is that there are no black and 60 yellow balls? If so, does this show that Levi’s theory needs a measure of epistemic reliability,  $\rho$  – needs higher order probabilities? Perhaps not to solve the present problem, but with an ounce of creativity the degree of indeterminacy can be readily multiplied.

Plainly, it is not much of an argument to say that Levi's "solution" to Ellsberg's paradox feels somewhat counterintuitive. Our philosophical intuitions simply diverge. What is a problem, however, is that Levi's theory violates one of the fundamental ordering principles of rational choice, and it is this violation that affords him the explanatory power. Luce and Raiffa's Axiom 7, the *principle of independence of irrelevant alternatives*, says that if a decision alternative is suboptimal in a decision situation it should not be possible to make it optimal by adding new alternatives. Luce and Raiffa illustrate this with a charming example. "Doctor: Well, Nurse, that's the evidence. Since I must decide whether or not he is tubercular, I'll diagnose tubercular. Nurse: But, Doctor, you do not have to decide one way or the other, you can say you are undecided. Doctor: That's true, isn't it? In that case, mark him not tubercular" (Luce and Raiffa 1957, pp. 288–289). Levi's theory does not satisfy this condition, but the MmEU criterion does.

The following decision problem (Gärdenfors and Sahlin 1982; Sahlin 1985) demonstrates the problem. Imagine a decision matrix with two alternatives, two states and consequently four possible outcomes. If  $a_1$  is chosen and state 1 ( $s_1$ ) obtains the outcome is 0 (let us say, utilities) and if state 2 ( $s_2$ ) obtains the outcome is also 0.  $a_2$  gives 11 and -9, respectively. Assume that Levi's set  $B_{X,t}$  consists of all convex combinations of the two probability distributions  $P_1$  and  $P_2$ , defined by  $P_1(s_1) = 0.4$  and  $P_1(s_2) = 0.6$  and  $P_2(s_1) = 0.6$  and  $P_2(s_2) = 0.4$ . Both  $a_1$  and  $a_2$  are E-admissible, but  $a_1$  the only admissible alternative because of its better security level, it is security-optimal. However, if we add a new alternative,  $a_3$ , which gives -10 if  $s_1$  obtains and 12 if  $s_2$  obtains,  $a_2$  is all of a sudden rendered optimal by Levi's theory.

This simple three-by-two matrix shows another thing. Assume that in this situation Levi's  $B_{X,t}$  contains the same set of distributions as Gärdenfors and Sahlin's  $P/p_0$ . Levi's decision rules recommend  $a_2$ , but MmEU recommends  $a_1$ . Which of the two theories should the decision maker use as the basis of a decision? But the decision maker does not only have these two theories to choose between. Henry Kyburg (1983) has suggested a theory to be used in situations where the quality and quantity of information leads to unreliable probability estimates. His decision principle says that the decision maker ought to reject any choice  $c$  for which there is an alternative whose minimum expected utility exceeds the maximum expected utility of  $c$ . Kyburg's theory tells us that we have no reason to choose one of the alternatives instead of another. Just for fun, let us now invent and apply a "new," fourth generalized "theory" with renewed decision rules. It is not hugely plausible to do this, but why not replace MmEU with a maxi–max rule, MMEU? Such an amendment would deliver a theory that recommends  $a_3$ , an alternative recommended by none of the three theories so far considered.

In an emergency ward decisions are made constantly, under great pressure of time and in circumstances that are stressful in other ways, too. There is not always time for reflective decision-making: there is a risk that affects influence the hard choices too much. Sometimes the quality and quantity of information is such that the probabilities are unreliable. Also the values can be more or less indeterminate – sufficiently unreliable and indeterminate to make the classical theories inapplicable. In situations like this a decision support system could aid the doctor's decision-making. Assume that such a tool has been developed. The developers, knowing that there are competing generalized decision theories, have programmed the machine to use several theories and present their converging or diverging recommendations. A patient is in a critical condition. The doctor feeds the decision tool with the information available, presses "run," and waits for advice. Unfortunately the computer gives not one, but several recommendations. Theory 1 recommends that a complicated and dangerous operation

is performed immediately. Theory 2 recommends new tests and that the patient be put under observation. Theory 3 regrets that, in this type of situation, it cannot give one and only one recommendation, since there are too many equally good options. What shall the doctor do now? It seems that to settle on just one of the conflicting recommendations, he or she will have to know as much about the theories as those who developed them. That may be asking too much of a decision maker.

Savage was aiming at a complete normative theory. So was Ramsey – at least, if we should put our trust in Rational Man or *Homo economicus* (two ideals in one). Strait is the gate and narrow is the way, but not if we opt for one of the generalized theories. Which theory should we choose? Is there a way to tell which is preferable? If there were a definite answer to this question, would not that imply that we knew what the ideal generalized theory would look like? But obviously there is no such theory, since this type of theory always gives up one or other of the commandments of classical theory. At best we can use several theories and hope for converging recommendations. That is a good thing. Or we could investigate and try to say under what circumstances one theory performs better than another, but this too seems a hopeless task.

We have seen that the classical theories of rational decision-making are not useful if probabilities and values are unreliable. We have also seen that the attempts to develop more general decision theories, designed to deal with this type of higher order uncertainty, all have their own problems. Leaving the straight and narrow road traveled by the classical decision theories leads straight down to too warm a place for any decision theoretician. On the flanks of the choice between giving up ordering or giving up independence lie Scylla and Charybdis.

## Epistemic Risk

---

Keynes (1933) once wrote: “When he [Ramsey] did descend from his accustomed stony heights he still lived without effort in a rarer atmosphere than most economists care to breathe.” Ramsey’s theories are indeed crystal-clear, complicated masterpieces; they do not make for easy reading, and they involve theoretical assumptions dissociating them from the types of problem we – you and I, or society at large – have to discuss and solve. We have seen that the traditional theories of rational decision-making are perfect tools for action if there is no epistemic uncertainty or value uncertainty. But, examining some contemporary risk debates (e.g., GMOs, nanotechnology, stem cell research, electromagnetic fields), we find that the evaluation of epistemic uncertainty seems as problematic as the assessment of the disutility of potential negative outcomes. The epistemic uncertainty causes epistemic risks, making it hard to identify, assess, and manage the outcome risks, i.e., the different negative things that might or might not happen as the result of our decisions.

A good example is nano-safety. Nanotechnology offers tremendous technological breakthroughs, but its promises come with a risk. It is impossible to say what the risk of nanotechnology is. There is no such (single, overall) risk. What can be assessed is the risk of individual innovations, inventions, tools and methods. At the same time, this field of research moves almost at the speed of light. This means that the systematization described below could well have been overtaken, or entrenched for that matter, by results published since this article was sent to the editors. I will therefore say “we believe” instead of “we know.” And our beliefs are sometimes based on a single article – sometimes supported by collateral evidence.

There are many types of nanoparticle. Fullerenes are molecules composed of carbon. Buckyballs are spherical fullerenes and nanotubes are cylindrical in shape. We also have, for

example, nanobuds (the combination of tubes and buckyballs – tubes with a verruca) and nanowires.

Let us briefly look at some of the health effects we believe that fullerenes can have. But also indicate what type of knowledge we lack. (A good survey of our present toxicological knowledge can be found in Ostiguy et al. (2008). See also Hermerén 2007 and 2008.)

*Nanotubes.* There are studies indicating, for example, that nanotubes can pass through the cellular membrane and that they can accumulate in the cell and end up in the cell nucleus. Other studies indicate that nanotubes have an effect on the respiratory system and that they can cause mechanical blockage of the airway (studies on rats). We believe that nanotubes are toxic, and that their toxicity depends on, among other things, size. There is also a study indicating that single-stranded DNA interacts with the nanotube (Zheng et al. 2003). If so, does this have an effect on transcription or methylation processes?

However, we do not know if nanotubes have an effect on, for example, the nervous system, or the liver, or whether they have developmental implications.

The problem is that today the quality and quantity of the information we have and knowledge we believe we have is not very robust. There are considerable epistemic uncertainties.

*Buckyballs.* There are studies indicating that buckyballs, spherical carbon nanoparticles, can affect metabolism. Toxicity studies have shown kidney effects resulting in a decrease in body weight and dwarfed vital organs. And several studies indicate that this type of particle can have an undesired effect on the respiratory system.

However, so far we do not know if these particles have an effect on the gastrointestinal system, or on the reproductive system or reproductive outcomes. Nor do we know whether, and if so, under what precise circumstances, they are carcinogenic.

Another complication is that most studies have been done on so-called  $C_{60}$  balls/molecules (a truncated icosahedron with a diameter of 1.1 nm). This buckyball can be found in, for example, soot. However, there are also buckyballs with 70, 72, . . . 100 carbon atoms. Again, there are Boron buckyballs – for example,  $B_{80}$ . The studies on  $C_{60}$  do not give us a robust picture of the possible health effects of other balls.

The problem is that today the quality and quantity of the information we have is not very robust. There is considerable epistemic uncertainty.

*Nanowires.* There are metallic, semiconducting, and insulating nanowires. Nanowires are one-dimensional and have fascinating properties that we do not find in three-dimensional materials.

There is an expectation that we will be able to press these wires into all sorts of service, but much is still at an experimental stage. One idea is that they will become the nuts and bolts of the next generation of “computers.” By following the movements in the surroundings of nanowires, nanoscientists have shown that these wires can generate electricity (e.g., see Xu et al. 2010). It takes but a scrap of imagination to see that nanowires will, potentially, be used for deep brain stimulation – for example, in the battle against Parkinson’s disease – or used in the treatment of patients with various types of heart condition.

To date few studies have looked at the health effects of nanowires. The epistemic uncertainty is, if not complete, considerable. Eriksson Linsmeier et al. (2009) studied brain-tissue response to nanowire implantations (in rats). They found that nanowire implantation induced both an astrocyte reaction and a microglial response, but that both phenomena declined over time. The authors found “no significant difference in the neuronal fraction for the nanowire-implanted animals compared to controls at all time points” but that that some nanowires were

able to pass the blood-brain barrier and leave the brain. They concluded that nanowires have short- and long-term effects which “require precaution and further investigation.” Since the types of nanowire we construct in labs are not found in nature, it is questionable whether we, or the environment, are adapted to adjust to this type of particle. Abundant reasons make it obvious that we need to improve our epistemic state.

Once again, the problem is that today the quality and quantity of the information we have is not very robust. There is considerable epistemic uncertainty.

To sum up. At present it seems that we know with differing degrees of epistemic certainty that nanoparticles can pass through various protective barriers, spread throughout the body, and accumulate in organs (e.g., the lungs and brain) and cells. We also believe we know that factors such as specific surface, surface properties, and number of particles have an effect on such things as toxicity. A worry is the potential accumulation of nanoparticles. This may or may not result in new properties – over time larger particles with new properties build up, particles that may or may not be harmful. We simply do not know.

There are four paradigmatic types of decision. The complete picture is, of course, greatly more complicated than a simple four-by-four matrix.

In *Type 1* situations the decision maker has extensive knowledge and information, expressed in terms of precise probability estimates. He has also clear and distinct preferences and values.

*An example* of a Type 1 situation: In *The Foundations of Statistics* Savage (1954/1972) asks us to consider a situation in which your wife or husband has just broken five good eggs into a bowl. A sixth egg lies beside the bowl and you have to decide what to do with it. You can break it into the bowl, break it onto a plate for inspection, or throw it away without inspection. If you know your eggs you know the probability that the sixth egg is also good. There is no epistemic uncertainty. And clearly a six-egg omelet is better than a five-egg omelet and the chore of washing up a plate, which in turn is better than five destroyed eggs and no omelet at all.

In *Type 2* situations the quality and quantity of information is poor, and it is difficult to represent the underlying uncertainty in terms of probability. On the other hand, the decision problem is one in connection with which the decision maker still has clear and distinct preferences and values: he knows what he wants and desires.

*An example* of a Type 2 situation: Many of the most well-known, contemporary risk issues fall within this category: GMOs, stem cell research and therapies, nanotechnology, and synthetic biology. We all want the good things that the new technologies promise. But we do not want the risks, and given our present state of knowledge it is also hard to say with precision what, and how great, the risks are. The possible risks of nanotubes and buckyballs are but one example. There are others. A group of scientists, the JCVI team, has recently shown how a small bacterial genome can be synthesized, and how this synthetic genome can be put into a cell to create a cell controlled completely by the synthetic genome (Gibson et al. 2010). As they point out, instead of (just) reading the book of nature we can now write it. Little imagination is needed to see the good things this technology could, potentially, offer. One idea is that in the future we might be able to create energy-producing bacteria, thereby dealing with a number of environmental problems. Another is that we might put ourselves in a position to create far more efficient vaccines for numerous diseases. However, there are risks – are there not? – to both the environment and us as individuals. But what are the risks? How great are they? What does our present epistemic state look like?

In *Type 3* situations the quality and quantity of information is good – good enough to assess precise probabilities. However, the decision maker lacks harmonious, clear and distinct preferences and values. Perhaps his appetites are out of keeping with his valuations.

An example of a Type 3 situation: ICDs (Internal Cardiac Defibrillators) are implants monitoring the patient's heart rhythm. If life threatening arrhythmias are detected by the ICD device it gives the patient a stinker of an electric shock. Before the device is implanted the patient is informed about the risks and benefits. A problem is that the device can be triggered when the patient has no arrhythmia. The probabilities relevant to a decision are known with reasonable precision, but they are known at type, not token, level. This is a decision-situation where our preferences and utilities are blurred or all but indistinct. We lack experience. An experience we hope we never will have. A special problem is that we have to choose between a quick and painless death, on the one hand, and a long drawn-out and maybe not especially painless ending, on the other.

In *Type 4* situations both information and preferences are unfixed or unreliable.

An example of a Type 4 situation: Imagine a 3-year-old child with high-risk neuroblastoma (Castor and Sahlin 2008). She has been treated according to the standard protocols used by pediatricians: she has been given hematopoietic stem cell transplantation, for example – a treatment not without complications and considerable suffering. But not all children respond to stem cell treatment, and let us assume this is a case of recurring neuroblastoma. What are the odds that a hematopoietic stem cell transplantation that failed to succeed the first time will work at the third or fourth attempt? With the number of refractory instances, the relevant probabilities become harder and harder, if not impossible, to estimate; and as our epistemic state deteriorates, our desires and values become more and more unstable.

Type 1 situations are readily detected in the paradigmatic cases with which classical theories of rational choice and decision-making deal. But the traditional theories are ill equipped to handle the three other types of situation. We have seen that in situations of Types 2, 3 and 4 the traditional theories are no guides to action. Here we need theories that help us to represent unreliable or indeterminate beliefs and imprecise values. We must introduce more complex, but also more complete, decision procedures. “Maximize expected utility,” the mantra of the classical theories, is, for simple mathematical reasons, no longer an available option. This means that we must use one or more of the generalized theories as a basis for action. Ideally these will deliver the same recommendation.

Which of the four situation-types best describes nanoscience and the risks of nanotechnology? The answer is: all four. We have to address the individual risk problems one by one, not together in an undifferentiated lump. Furthermore, as our knowledge improves we have to

		Epistemic uncertainty	
		Low	High
Value uncertainty	Low	Type 1	Type 2
	High	Type 3	Type 4

Fig. 18.1  
Types of decision

revalue the risks. What to begin with was a Type 4 situation can, with the arrival of new knowledge and information, turn into a Type 3 situation and end up as a Type 1 problem.

Examples involving nuclear power plants, BSE, GMOs, electro-magnetic fields, and nanotechnology have taught us that various kinds of factor create epistemic uncertainty. Good risk analysis requires careful inspection of our present epistemic state. It is simply not enough to identify and evaluate outcome risks, i.e., the negative consequences of our actions. An estimation and evaluation of current levels of ignorance is crucial. We must know what type of knowledge we have and the value situation we are in (Type 1, 2, 3 or 4).

If we cannot use the traditional methods for rational decision-making and risk analysis, what shall we do? The generalized theories discussed above are good tools and help a lot, but as we have seen they are less than perfect. What we should do, I think, is promote a Socratic approach to risk analysis. What does this mean?

In Plato's *Apology* Socrates declares: "Probably neither of us knows anything really worth knowing: but whereas this man imagines he knows, without really knowing, I, knowing nothing, do not even suppose I know. On this one point, at any rate, I appear to be a little wiser than he, because I do not even think I know things about which I know nothing" (Blakeney 1929, pp. 67–68).

Socrates tells us that it is important that we monitor the epistemic uncertainties we have. This means that we should not let statistics and traditional methods of decision analysis force us to pretend that we are in Type 1 situations when in fact we are glued in the opposite corner dealing with Type 4 decisions. Instead we shall give an as-frank-as-possible picture of the epistemic situation – trying our best to say what is known, describing the stability of our knowledge, and honestly conceding what we do not know. This, I believe, is the only way to avoid the prospect of the new technologies becoming stigmatized, and loss of public trust. The bottom line is that we have a moral obligation to give an honest, complete description of our epistemic state.

But how is this goal to be achieved? We must identify factors that produce epistemic uncertainty. A complete list of factors is impossible to give, but here are some important and well-known factors producing epistemic uncertainty and value uncertainty (Sahlin 1992; Sahlin and Persson 1994; Sahlin et al. 2010; Vareman et al. 2011).

*The unreliable research process.* Research is a mechanism which, off-and-on, gives us incorrect or indeterminate results. Sometimes the machinery works flawlessly, sometimes chance has an unfavorable effect on the results, and sometimes the investigation does not work at all. To accept the results without asking whether or no they are the result of a working mechanism is to wink at all forms of epistemic uncertainty. A Socratic approach to risk analysis demands of us that we carefully scrutinize the strength of the different pieces of evidence we have, that we assess their evidentiary value and find out if they corroborate, or conflict with, each other.

*The fact of irrationality.* Contemporary psychologists have taught us a great deal about the way we perceive risks, and about the way affects and emotions influence our behavior. Their research has shown that, as decision makers, as risk-assessors, and as risk-controllers, we are shortsighted, one-eyed, and prone to serious errors of refraction. We generate too few, and too narrow, hypotheses. We gather information, or evidence, in favor of our guesses that is too narrow, readily available, and skewed in favor of preferred beliefs. Once we have a pet hypothesis, we look for confirmatory material, neglecting countervailing evidence. We are simply not rational – not in the way our theories of rationality (logic, probability and decision-making) assume, at any rate. This is an alarming fact, considering the serious risk-assessment and risk-management tasks that lie ahead of us.

The fact of irrationality then pushes us in the direction of epistemic uncertainty. It sends us into Type 2 and Type 4 situations. This must be avoided, and it can only be avoided if we take a Socratic approach to what we are doing.

*The choice and lack of theories.* Physics is the science of prediction. But sometimes even physicists run into problems. The n-body problem is still elusive. Many differential equations are too hard to solve, and if we seek to replace them with one we can solve we might end up with a rather bad approximation or model. Redhead (1980) gives a simple, powerful and illuminating example. Suppose our theory tells us that  $(1 - x)dy/dx = \cos \lambda x + y$ . And assume we study the theory using an approximation, in terms of a power series of  $x$ , and a model in which  $\cos \lambda x$  is set to 1. As Redhead points out, both the approximation and the model will mislead us, for in different ways they introduce epistemic uncertainty. That theories are permeated by epistemic uncertainty is shown by, for example, the vacuum catastrophe. The calculations undertaken by Quantum field theory missed the target by an order of 107 (see Bax 2004). Some of the phenomena the nanoscientist can observe are also hard work for the theoretician. This, of course, is a problem. We want to be able to say what will happen in advance, not be surprised by experiments which in turn are hard to underpin with sound theoretical explanations. And prediction becomes even more of a challenge when we merge physics and biology. Can we predict anything here, or do we have to search for knowledge by carefully designed experiment? The science is one of small steps.

Traditionally, toxic risks are understood in terms of dose-response ratios. However, nanoparticles do not behave like the chemicals with which we are more familiar. For them, dose, understood in terms of concentration and mass, is no longer an applicable model (Ostiguy et al. 2008). Factors like specific surface, dimension, number of particles, and surface properties are often far more important, and it must not be forgotten that the properties of the nanoparticles can change or become totally different when the particle size decreases or increases, or when particles lump or interact. For similar reasons it seems that in the case of nanoparticles the step from in vitro to in vivo might be bigger than usual. We need new models. And, adding to the epistemic uncertainty, we do not know much about species differences. Today's deficiency in models, theories and data simply yields a considerable amount of epistemic uncertainty – an uncertainty that we should take into account when making decisions, but one that should also arouse our curiosity and inspire new research.

*Unrealizable research.* We might get caught in situations where, for moral or practical reasons, it is difficult to carry out controlled experimental studies. As a result we might have to rely on indirect evidence rather than solid, direct empirical evidence.

The best way to discover the effects of toxic substances on humans is to conduct experiments on humans, not animals. But that is often impossible, because the tests would be unethical, so we end up relying on animal experimentation. Our moral commitments, then, create uncertainty. They produce epistemic risk. But that is a fact that we have to accept.

There are also practical problems. In testing toxic substances, one works with three dose groups: a control group (zero dose) and two groups of animals exposed to higher or lower doses of the substance. In a 28-day dose toxicity test, 5 animals of each sex and dose group are used, i.e., 30 animals in total. To obtain significant results with groups of that size, the experiment must be carried out at high dose levels well above those we are normally exposed to. In the light of this type of experiment, it is difficult to say what increase in risk can be expected at the normal dose level. In addition to statistical limitations and the limitations of the experimental design, there is the problem of extrapolation. The hamster, for example, is 10,000

times less sensitive to some toxins than the guinea pig, and man is neither a hamster nor a guinea pig.

To return to nanowires, the complex properties these have will raise new questions about the type of knowledge we can obtain. They also favor a Socratic approach to risk analysis – i.e., the idea that we should avoid thinking we know things about which we know nothing (and maybe never can know robustly).

*Time, a problem.* Time causes a particular problem when it comes to risk assessment and risk management. What will the environmental effects of nanoparticles be in the long run? How do the particles affect human beings over the longer term? Assume that it is the accumulation of particles that is the biggest risk. Assume that the effect of the particles first becomes visible after several decays. In the case of nanoscience and nanotechnology it is today (but not necessarily tomorrow), and for obvious reasons, it is difficult, if not impossible, to say with any precision what will happen in the long run. It is also, for as obvious reasons, difficult to study this type of problem or phenomenon. The further possibility that the traditional dose-response model is not applicable might add to the difficulties here.

Why is it significant to know what we do not know – especially when it comes to risk-assessment and risk management? Why is this the best way to do risk analysis?

First, we cannot be rational if we do not take our complete state of knowledge into account.

Second, the Precautionary Principle says that “when an activity raises threats of harm to human health or the environment, precautionary measures should be taken even if some cause and effect relationships are not fully established scientifically.” There are several problems with a principle of this kind (Peterson 2006, 2007). It can, for example, come in conflict with some of the basic principles (axioms) of rational decision-making, and if it is not applied with care it is also capable of paralyzing research. A Socratic approach allows for rational decision-making when the epistemic uncertainty is considerable – e.g., using the tools and techniques of the generalized theories of rational decision-making (presented and discussed above) (see Vareman 2011).

Third, psychological studies have taught us that if its attendant uncertainties and risks are not handled with care an entire research area can become stigmatized. The public feel they can no longer trust the scientists, and trust is essential when it comes to risk-communication (Slovic 1997). I contend that a Socratic approach builds trust and blocks stigmatization.

Fourth, knowing what you do not know is the spark of creativity. If we have indications that a particular particle is harmful, but we do not really know if it is, why should that fact or feature of our present epistemic state make us gloomy? It should make us curious, inspire us to look for the underlying mechanisms, encourage us to seek out new knowledge. More than once a lack of knowledge and connected efforts to avoid epistemic uncertainty have been rewarded with a Nobel Prize.

## Further Research

---

Two suggestions for further research conclude this paper.

Saari (1994) has shown how voting theory can be explained with the tools of classical geometry. He shows that with a sufficiently powerful mathematical framework competing ideas and methods can be understood and unified, controversies can be resolved, and famous theorems become predictable spin-offs. Is it possible to do something similar for classical and generalized theories of rational decision-making? Can we provide a geometry or topology of

decision-making? We want to develop a higher-level theory that can be used to evaluate existing theories and in understanding decision rules – a theory that tells us, for example, when and under what circumstances a generalized theory works, but also when it does not work, and then why.

Risk analysis has focused too much on outcomes, on accidents, catastrophes, mishaps, and harm. The examples presented above show that epistemic factors are as important, sometimes far more important. It is therefore essential that epistemic uncertainty and epistemic risk be researched. To make good decisions we need to understand how epistemic uncertainty is produced and in what way it influences our decision-making and risk-taking. The problem of epistemic risk needs to be addressed both by the theoreticians and those doing empirical research.

## Acknowledgments

The author wishes to thank Naomi Halas, Johannes Persson, Christelle Prinz, Bengt E Y Svensson, and Niklas Vareman for valuable comments and suggestions. A special thanks to Pascal Engel and Kevin Mulligan for inspiration and time for reflection.

## References

- Allais M (1953) Le comportement de l'homme rationnel devant le risque: critique des postulats et axioms de l'école Americaine. *Econometrica* 21:503–546
- Anscombe FJ, Aumann RJ (1963) A definition of subjective probability. *Ann Math Stat* 34:199–205
- Bax Ph (2004) The cosmological constant problem. *Contemp Phys* 45:227–236
- Blakeney EH (ed) (1929) The apology of Socrates. The Scholartis Press, London, pp 67–68
- Bolker ED (1967) A simultaneous axiomatization of utility and subjective probability. *Philos Sci* 34:333–340
- Bränmark J, Sahlin N-E (2010) Ethical theory and philosophy of risk: first thoughts. *J Risk Res* 13:149–161
- Brehmer B, Sahlin N-E (1994) Future risks and risk management. Kluwer, Boston
- Castor A, Sahlin N-E (2008) Mycket svåra beslut. In: Persson J, Sahlin N-E (eds) *Risk & risici. Nya Doxa*, Riga, pp 231–248
- Davidson D, Suppes P, Siegel S (1957) Decision making: an experimental approach. Stanford University Press, Stanford
- de Finetti B (1937) La prévision: ses lois logiques, ses sources subjectives. *Ann Inst Henri Poincaré* 7:1–68. Presses universitaires de France, Paris. Theory of probability, vol I, 1974 and II, 1975. Wiley, New York
- Debreu G (1959) Cardinal utility for even-chance mixtures of pairs of sure prospects. *Rev Econ Stud* 26: 174–177
- Ellsberg D (1961) Risk, ambiguity and the savage axioms. *Q J Econ* 75:643–669. Reprinted as chapter 13 in
- Gärdenfors P, Sahlin N-E (1988) *Decision, probability and utility*. Cambridge University Press, Cambridge
- Eriksson Linsmeier C, Prinz CN, Pettersson LME, Caroff P, Samuelson L, Schouenborg J, Montelius L, Danielsen N (2009) Nanowire biocompatibility in the brain looking for a needle in a 3D stack. *Nano lett* 9:4184–4190
- Fishburn PC (1981) Subjective expected utility: a review of normative theories. *Theory Decis* 13:139–199
- Gärdenfors P, Sahlin N-E (1982) Unreliable probabilities, risk taking, and decision making. *Synthese* 53:361–386. Reprinted in Gärdenfors P, Sahlin N-E (1988), pp 331–334
- Gärdenfors P, Sahlin N-E (1983) Decision making with unreliable probabilities. *Br J Math Stat Psychol* 36:240–251
- Gärdenfors P, Sahlin N-E (eds) (1988) *Decision, probability, and utility: selected readings*. Cambridge University Press, Cambridge
- Gibson DG et al (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329(5987):52–56
- Hermerén G (2007) Challenges in the evaluation of nano-scale research: ethical aspects. *NanoEthics* 1:223–237
- Hermerén G (2008) Ethical aspects of nanomedicine: a condensed version of the EGE opinion 21. In: Allhoff F, Lin P (eds) *Nanotechnology and society*:

- current and emerging ethical issues. Springer, Dordrecht, pp 187–206
- Jeffrey RC (1965) The logic of decision. McGraw-Hill, New York. (Second revised edition 1983) University of Chicago Press, Chicago
- Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. *Econometrica* 47: 263–291
- Keynes JM (1921/1957) A treatise on probability. Macmillan, London
- Keynes JM (1933) Frank Plumpton Ramsey. Essays in biography. The Norton Library, New York
- Kyburg H (1983) Rational belief. *Behav Brain Sci* 6: 231–273
- Levi I (1980) The enterprise of knowledge. MIT Press, Cambridge, MA
- Levi I (1986) Hard choices. Cambridge University Press, Cambridge
- Levi I (1988) On indeterminate probability. In: Gärdenfors P, Sahlin N-E (eds) Decision, probability, and utility: selected readings. Cambridge University Press, Cambridge, pp 286–312
- Luce RD, Raiffa H (1957) Games and decisions: introduction and critical survey. Wiley, New York
- Ostiguy C, Soucy B, Lapointe G, Woods C, Ménard L, Trottier M (2008) Health effects of nanoparticles, 2nd edn. IRSST: Chemical substances and biological agents: Studies and research projects report R: 589
- Peterson M (2006) The precautionary principle is incoherent. *Risk Anal* 26(3):595–601
- Peterson M (2007) Should the precautionary principle guide our actions or our beliefs? *J Med Ethics* 33:5–10
- Ramsey FP (1926/1990) Truth and probability. In: Mellor DH (ed) Philosophical papers. Cambridge University Press, Cambridge/New York, pp 52–94
- Ramsey FP (1990) Weight or the value of knowledge, Sahlin N-E (ed). *Br J Philos Sci* 41:1–3
- Ramsey FP (1991) Notes on philosophy, probability and mathematics, Galavotti MC (ed) Bibliopolis, Napoli
- Redhead M (1980) Models in physics. *Br J Philos Sci* 31:145–163
- Saari D (1994) Geometry of voting. Springer, Berlin
- Sahlin N-E (1983) On second order probabilities and the notion of epistemic risk. In: Stigum BP, Wenstøp FD (eds) Foundations of utility and risk theory with applications. Reidel, Dordrecht, pp 95–104
- Sahlin N-E (1985) Three decision rules for generalized probability representations. *Behav Brain Sci* 4:751–753
- Sahlin N-E (1988) The significance of empirical evidence for developments in the foundations of decision theory. In: Batens D, van Bendegem JP (eds) Theory and experiment. Reidel, Dordrecht, pp 103–121
- Sahlin N-E (1990) The philosophy of F. P. Ramsey. Cambridge University Press, Cambridge
- Sahlin N-E (1991) Baconian inductivism in research on human decision making. *Theory Psychol* 4:431–450
- Sahlin N-E (1992) Kunskapsrisk, utfallsrisk och moraliskt instabila beslut. In: Risk, bioteknologi och etik, Nordiske Seminar- og Arbejdsrapporter: 503. Nordisk Ministerråd, Copenhagen
- Sahlin N-E (1993) On higher order beliefs. In: Dubucs JP (ed) Philosophy of probability. Kluwer, Boston, pp 13–34
- Sahlin N-E, Brännmark J (2011, forthcoming) How can we be moral when we are so irrational?
- Sahlin N-E, Persson J (1994) Epistemic risk: the significance of knowing what one does not know. In: Brehmer B, Sahlin N-E (eds) Future risks and risk management. Kluwer, Boston, pp 37–62
- Sahlin N-E, Vareman N (2008) Three types of decision theory. In: Galavotti MC, Scazzieri R, Suppes P (eds) Reasoning, rationality and probability. CSLI, Stanford, pp 37–59
- Sahlin N-E, Persson J, Vareman N (2010) Unruhe und Ungewissheit – stem cells and risks. In: Hermerén G, Hug K (eds) Translational stem cell research: issues beyond the debated on the moral status of the human embryo. Springer, Totowa
- Savage LJ (1954) The foundations of statistics. Wiley, New York. (Second revised edition 1972) Dover, New York
- Schervish M, Seidenfeld J, Kadane J (1990) State-dependent utilities. *J Am Stat Assoc* 85(411):840–847
- Seidenfeld T (1988) Decision theory without ‘independence’ or without ‘ordering’: what is the difference? *Econ Philos* 4:267–315
- Slovic P (1997) Trust, emotion, sex, politics, and science: surveying the risk-assessment battlefield. In: Bazerman MH, Messick DM, Tenbrunsel AE, Wade-Benzoni KA (eds) Environment, ethics, and behavior. New Lexington, San Francisco, pp 277–313. Also in (1999) *Risk Anal* 19(4):689–701
- Suppes P (1956) The role of subjective probability and utility in decision-making. In: Neyman I (ed) Proceedings from third Berkley symposium on mathematics, statistics and probability. University of California Press, Berkeley, pp 61–73
- Vareman N (2011, forthcoming) The normative basis of risk analysis
- Vareman N, Persson J, Sahlin N-E (2011, forthcoming) Conceptions of epistemic risk
- von Neumann J, Morgenstern O (1944) Theory of games and economic behavior. Princeton University Press, Princeton. (Second edition 1953) Wiley, New York
- Xu S et al (2010) Self-powered nanowire devices. *Nat Nanotechnol* 5:366–373
- Zheng M, Jagota A, Strano M, Santos A, Barone P (2003) Structure-based carbon nanotube sorting by sequence-dependent DNA assembly. *Science* 302: 1545–1548

# 19 Paradoxes of Rational Choice Theory

Till Grüne-Yanoff

University of Helsinki, Helsinki, Finland

<i>Introduction</i> .....	500
<i>Rational Choice Theory</i> .....	500
<i>Normative Validity and the Role of Paradox in RCT</i> .....	502
<i>The Notion of Paradox</i> .....	504
<i>The Paradoxes</i> .....	505
Preferences .....	505
Belief .....	508
Expected Utility .....	508
Strategic Interaction .....	511
<i>Further Research</i> .....	514

**Abstract:** Rational choice theory (RCT) is beset with paradoxes. Why, then, bother with a theory that raises numerous counterexamples, contradictions, and a seemingly endless stream of mutually conflicting remedies? In contrast to this impression, I argue in this chapter that RCT paradoxes play much more productive roles. Eight paradoxes are described in detail, some of their proposed solutions are sketched out, and they are classified according to the kind of paradox they pose. At their example I argue that RCT paradoxes, rather than providing evidence for straightforward rejections of the theory, play important roles in education and in normative research.

## Introduction

---

Rational choice theory (RCT) is beset with paradoxes. Why, then, bother with a theory that raises numerous counterexamples, contradictions and a seemingly endless stream of mutually conflicting remedies? That, at least, may be the impression of a novice student of RCT. In contrast, I argue in this chapter, RCT paradoxes play a much more productive role. Rather than suggesting straightforward rejections of the theory, or repellents to any newcomer, paradoxes play important roles in education and in normative research.

RCT has a clear normative function: it offers tools for judging how people *ought* to form their preferences – and by extension, how they ought to choose. A major problem is that there is no hard basis against which to test normative theoretical claims – one cannot seek to falsify such a theory with controlled experiments. Instead, researchers have to rely on normative intuitions about assumptions and conclusions, and use theory to check whether these intuitions can be held consistently. This is where RCT paradoxes play a crucial role: they elicit normative intuitions that pitch RCT assumptions and conclusions against each other. If a paradox leads to a revision of the theory, it serves a research purpose. If it leads to a better understanding of the assumptions and their conclusions, it serves an educational purpose. Thus, many RCT paradoxes have proved, and continue, to be productive.

The chapter continues with a brief overview of RCT in the first section, recalling the normative claims it really makes. The second section discusses how its normative validity can be examined, and the roles paradoxes play in that. The third section offers a classification of different kinds of paradoxes. Eight selected paradoxes are surveyed in the fourth section, sub-sectioned into paradoxes of preference, belief, expected utility, and strategic interaction. Each one is explained, some of its proposed solutions are sketched out, and it is classified according to the scheme proposed in the fifth section. Section ➔ [Further Research](#) concludes the chapter.

## Rational Choice Theory

---

RCT is the dominant theoretical approach in microeconomics (although economists rarely use the term “rational choice theory”). It is also widely used in other social-science disciplines, in particular political science. In this context, the term rational choice theory is often associated with the notion of economic “imperialism,” implying that its use extends economics methodology into their fields.

Explicit theories of rational economic choice were first developed in the late nineteenth century, and commonly linked the choice of an object to the increase in happiness an additional increment of this object would bring. Early neoclassical economists (e.g., William Stanley Jevons) held that agents make consumption choices so as to maximize their own happiness. In contrast, twentieth-century economists disassociated RCT and the notion of happiness: they presented rationality merely as maintaining a consistent ranking of alternatives. Such a ranking is commonly interpreted as agents' desires or values.

Having no foundation in an ultimate end, the notion of rationality is reduced to the consistent ranking of choice alternatives, the consistent derivation of this ranking from evaluations of possible outcomes, and a consistency of beliefs employed in this derivation. Thus, "rationality" explicated in RCT is considerably narrower and possibly sometimes at odds with colloquial or philosophical notions. In philosophical contexts it often includes judgments about ends, the prudent weighting of long-term versus short-term results, and insights into purportedly fundamental moral principles. Nothing of this sort is invoked in RCT, which simply claims that a rational person chooses actions in a manner consistent with his or her beliefs and evaluations. Accordingly, a person considered "rational" in this sense may believe that the moon is made of green cheese, may desire to waste his or her life, or may intend to bring widespread destruction.

At the core of RCT is a formal framework that (1) makes the notion of preference consistency precise and (2) offers formal proof that "maximizing one's utility" is identical to "choosing according to a consistent preference ranking." A brief sketch of this framework follows. (The framework presented here is based on von Neumann and Morgenstern 1947. Alternative formal frameworks are to be found in Savage 1954 and Jeffrey 1990.)

Let  $A = \{X_1, \dots, X_n\}$  be a set of alternatives. Alternatives are either pure prospects or lotteries. A pure prospect is a future event or state of the world that occurs with certainty. For example, when purchasing a sandwich from a well-known international restaurant chain I may expect certain taste experiences with near certainty. Lotteries, also called prospects under risk, are probability distributions over events or states. For example, when consuming "pick-your-own" mushrooms an agent faces the lottery  $(X_1, p; X_2, 1-p)$ , where  $X_1$  denotes the compound outcome (which has probability  $p$ ) of falling ill due to poisoning and  $X_2$  (with probability  $1-p$ ) the compound outcome of not doing so. More generally, a lottery  $X$  consists of a set of prospects  $X_1, \dots, X_n$  and assigned probabilities  $p_1, \dots, p_n$ , such that  $X = (X_1, p_1; \dots; X_n, p_n)$ . Obviously, the prospects  $X_1, \dots, X_n$  can be lotteries in themselves.

RCT takes preferences over actions to be evaluations of lotteries over action outcomes. Its main contribution is to specify the relationship between preferences over actions, and preferences as well as beliefs over the compound outcomes of the respective lottery. It does so by proving *representation theorems*. Such theorems show that, under certain conditions, all of an agent's preferences can be represented by a numerical function, the so-called utility function. Furthermore, the utility numbers of an action (i.e., lottery)  $X = (X_1, p_1; \dots; X_n, p_n)$  and its compound outcomes  $X_1, \dots, X_n$  are related to each other through the following principle:

$$u(X) = \sum_i p_i \times u(X_i) \quad (1)$$

In other words, the utility of a lottery is equal to the sum of the utilities of its compound outcomes, weighted by the probability with which each outcome comes about. This is an important result that significantly constrains the kind of preferences an agent can have. Of course, because the representation result is a formal proof, all the constraining information must already

be present in the theorem's assumptions. I will sketch the main features of these assumptions here. (For a detailed discussion, see the references in footnote 2. For more in-depth overviews, see textbooks such as Luce and Raiffa (1957), Mas-Colell et al. (1995, Chaps. 1 and 6) and Resnik (1987). Hargreaves Heap et al. (1992, pp. 3–26) give an introductory treatment.)

RCT assumes that, at any time, there is a fixed set of alternatives  $A = \{X_1, \dots, X_n\}$  for any agent. With respect to the agent's evaluation of these prospects, it assumes that agents can always say that they prefer one prospect to another or are indifferent between them. More specifically, it assumes that the agent has a preference ordering  $\succeq$  over  $A$ , which satisfies the following conditions. First, the ordering is assumed to be *complete*, that is,

$$\text{either } X_i \succeq X_j \text{ or } X_j \succeq X_i \text{ for all } X_i, X_j \in A. \quad (2)$$

Second, the ordering is assumed to be *transitive*, that is,

$$\text{if } X_i \succeq X_j \text{ and } X_j \succeq X_k, \text{ then also } X_i \succeq X_k \text{ for all } X_i, X_j, X_k \in A. \quad (3)$$

Completeness and transitivity together ensure that the agent has a so-called weak ordering over all prospects.

The second domain in which RCT makes consistency assumptions concerns beliefs. In particular, it assumes that each rational agent has a *coherent set of probabilistic beliefs*. Coherence here means that beliefs can be represented as probability distributions that satisfy certain properties. In particular, it is assumed that there is a probability function  $p$  over all elements of  $A$ , and that this function satisfies the following assumptions: first, for any  $X$ ,  $0 \leq p(X) \leq 1$ ; second, if  $X$  is certain, then  $p(X) = 1$ ; third, if two alternatives  $X$  and  $Y$  are mutually exclusive, then  $p(X \text{ or } Y) = p(X) + p(Y)$ ; finally, for any two alternatives  $X$  and  $Y$ ,  $p(X \text{ and } Y) = p(X) \times P(Y|X)$  – in other words the probability of the alternative “ $X$  and  $Y$ ” is identical to the probability of  $X$  multiplied by the probability of  $Y$  given that  $X$  is true.

The third domain in which rational choice theory makes consistency assumptions concerns preferences over lotteries. In particular, it assumes the *independence condition*. If a prospect  $X$  is preferred to a prospect  $Y$ , then a prospect that has  $X$  as one compound outcome with a probability  $p$  is preferred to a prospect that has  $Y$  as one compound with a probability  $p$  and is identical otherwise: that is, for all  $X, Y, Z$ : if  $X \succeq Y$  then  $(X, p; Z, 1-p) \succeq (Y, p; Z, 1-p)$ .

These assumptions (together with a few others that are not relevant here) imply that preferences over lottery prospects  $X = (X_1, p_1; \dots; X_m, p_m)$  are represented by a utility function such that for all  $X, Y$ :

$$X \succeq Y \Leftrightarrow \sum_i [p_i \times u(X_i)] \geq \sum_i [p_i \times u(Y_i)]. \quad (4)$$

This formal result has been given different interpretations. My focus in the following is on the *normative* interpretation of RCT.

## Normative Validity and the Role of Paradox in RCT

RCT is often interpreted as a theory of how people *ought* to form their preferences – and by extension, how they ought to choose (for a history of this approach, see Guala 2000). Although the normative content of the theory is limited to the norms of a consistent ranking of choice

alternatives, I showed in the previous section that this notion of consistency depends on a number of substantial axioms. This raises the question of the *normative validity* of these axioms: why ought people to choose in accordance with them?

Various attempts have been made to defend the normative validity of RCT and its axioms. The most prominent justifications are pragmatic: they seek to show that agents who fail to retain RCT-consistency will incur certain losses. Two well-known examples are the *money pump* and the *Dutch book* arguments (for more on this and other normative justifications, see Hansson and Grüne-Yanoff 2009).

Interpreted literally, neither the money pump nor the Dutch book is very convincing. An agent could simply refuse to accept money-pumping trades or Dutch-booking bets. Thus, rationality does not literally require one to be willing to wager in accordance with RCT. Defenders of pragmatic justifications may argue that money pumps or Dutch books reveal possible vulnerability from RCT-violations: a RCT violator might have an incentive to accept a money pump trade or a Dutch book bet, while a RCT-abider does not. However, even such a hypothetical interpretation is problematic. For example, one could deny that the situations considered are normatively relevant to actual preferences. Consequently, it could be argued that norms of preference consistency are primitive in the sense that they are not derived from anything, and in particular not from pragmatic considerations.

Instead, some argue that normative judgments arise directly through human intuition, guided by reflection. These judgments are grounded in characteristic human responses of an emotional or motivating kind. (Such a view does not presuppose a non-cognitivist account of normative judgment. At least on the epistemological level, even cognitivist theories of normativity are likely to appeal to something like natural human responses – no doubt refined by education and reason – to explain how we identify moral facts and evaluate moral claims.) While considerations such as the money pump or the Dutch book may elicit such intuitions, it would be misguided to assume that pragmatic considerations form their basis. Rather, normative intuitions themselves are basic, and form the basis of normative validity judgments of RCT, in this view.

So much the worse for the normative validity of RCT axioms, one might be tempted to reply. To be sure, our emotional or motivating responses to questions of preference consistency often differ and are contradictory. Hence, it seems to follow that any proposed set of axioms is nothing more than the expression of a subjective intuition, fuelled at best by positional or rhetorical power.

Defenders of a stronger validity claim may respond in at least two ways to this challenge. First, they may point out that normative intuitions are not merely claimed to be valid individually, but rather that RCT makes a claim about the normative validity of the whole set of assumptions *and all the results deduced from it*. For example, the maximization of expected utility is a consequence of standard RCT axioms, not an axiom itself. If one has doubts about the normative validity of this conclusion, one has to trace it back to these axioms, re-check their validity, and weight one's doubts in the conclusion against one's confidence in them. This view of normativity thus rests on the idea of a reflective equilibrium: we “test” various parts of our system of normative intuitions against the other intuitions we have made, looking for ways in which some of them support others, seeking coherence among the widest set, and revising and refining them at all levels when challenges to some arise from others (for more on the method of reflective equilibrium, see Daniels 2008).

Second, the defender may point out that normative intuitions are widely accepted only if they withstand being tested in a communally shared effort of “normative falsification” (Guala 2000). Savage described this effort as follows:

- ▶ In general, a person who has tentatively accepted a normative theory must conscientiously study situations in which the theory seems to lead him astray; he must decide for each by reflection – deduction will typically be of little relevance – whether to retain his initial impression of the situation or to accept the implications of the theory for it (Savage 1954, p. 102).

Theorists have to engage in thought experiments in order to elicit these normative intuitions – or “initial impressions,” as Savage calls them. Thereby they investigate their normative intuitions in as wide a scope of hypothetical situations as possible, either challenging or confirming particular normative judgments. At the end of this process they only use the intuitions that hold up against normative falsification to challenge the theory:

- ▶ If, after thorough deliberation, anyone maintains a pair of distinct preferences that are in conflict with the sure-thing principle, he must abandon, or modify, the principle; for that kind of discrepancy seems intolerable in a normative theory (Savage 1954, p. 102).

Normative falsification and reflective equilibrium thus go hand in hand: the former generates “corroborated” normative intuitions, and the latter weighs the importance of these intuitions against conflicting intuitions in the theory under scrutiny.

How, then, does one go about normative falsification? How are “situations” constructed in which one obtains “initial impressions” that conflict with the theory? This is where RCT paradoxes come into play. These paradoxes are exemplar narratives of situations that have posed problems for RCT, many of which have been discussed amongst experts for decades. Sometimes agreed-upon solutions exist, and the paradox is used only for pedagogical purposes – to increase understanding of the theory or to illustrate the process of thought experimentation. At other times, competing solutions are offered, some of which may threaten the current theory. In that case, RCT paradoxes constitute the laboratory equipment of ongoing decision-theoretical research.

## The Notion of Paradox

---

Philosophers have distinguished between two accounts of paradoxes. The *argumentative model*, proposed by Quine (1966) and Sainsbury (1988), defines a paradox as an argument that appears to lead from a seemingly true statement or group of statements to an apparent or real contradiction, or to a conclusion that defies intuition. To resolve a paradox, on this account, is to show either (1) that the conclusion, despite appearances, is true, that the argument is fallacious, or that some of the premises are false, or (2) to explain away the deceptive appearances. The *non-argumentative model*, proposed by Lycan (2010), defines a paradox as an inconsistent set of propositions, each of which is very plausible. To resolve a paradox under this account is to decide on some principled grounds which of the propositions to abandon.

Consequently, the argumentative model allows distinguishing different kinds of paradoxes. Quine divides them into three groups. A *veridical paradox* produces a conclusion that is valid, although it appears absurd. (Quine thought of paradoxes pertaining to the truth of deductive

statements. In contrast, the validity of decision-theoretic assumptions and conclusions concerns normative validity, which may or may not be reducible to truth. I will therefore use “validity” where Quine spoke of “truth.”) For example, the paradox of Frederic’s birthday in *The Pirates of Penzance* establishes the surprising fact that a 21-year-old would have had only five birthdays had he been born in a leap year on February 29.

A *falsidical paradox* establishes a result that is actually invalid due to a fallacy in the demonstration. DeMorgan’s invalid mathematical proof that  $1 = 2$  is a classic example, relying on a hidden division by zero.

Quine’s distinction here is not fine enough for the current purposes. A falsidical paradox in his terminology, so it seems, can be the result of two very different processes. A genuine falsidical paradox, I suggest, identifies the root of the invalidity of the conclusion in the invalidity of one or more of the assumptions. In contrast, what I call an *apparent paradox* establishes the root of the invalidity of the conclusion in the unsoundness of the argument.

A paradox that is in neither class may be an *antinomy*, which reaches a self-contradictory result by properly applying accepted ways of reasoning. Antinomies resist resolutions: the appearances cannot be explained away, nor can the conclusion be shown to be valid, some premises shown to be invalid, or the argument shown to be unsound. Antinomies, Quine says, “bring on the crisis in thought” (1966, p. 5). They show the need for drastic revision in our customary way of looking at things.

The non-argumentative account rejects Quine’s classification, pointing out his assumption of an intrinsic direction in the relationship between “assumptions” and “conclusions.” This, so Lycan argues, may give the wrong impression that certain kinds of paradoxes are to be solved in particular ways. Conversely, he points out that two theorists may disagree on whether a paradox is veridical or not: “one theorist may find the argument veridical while the other finds the ‘conclusion’s’ denial more plausible than one of the ‘premises’” (Lycan 2010, p. 3). In what follows, I will make use of Quine’s classification. Nevertheless, I stress – in agreement with Lycan – that it is only to be understood as an indicator of how the majority of theorists have sought resolution, not as a claim about the intrinsic nature of the paradox itself.

## The Paradoxes

---

Below I survey a selection of paradoxes that are currently relevant to RCT. By relevant here I mean that they challenge one or more of the RCT axioms that are currently in wide use. For more comprehensive literature on paradoxes, also in RCT, see Richmond and Sowden (1985), Diekmann and Mitter (1986), Sainsbury (1988), and Koons (1992).

The survey is structured according to the aspect of RCT under challenge. As it turns out, it is not always clear which axiom is being challenged. I therefore divide the subsections into paradoxes of preferences, belief, expected utility maximization, and strategic choice.

### Preferences

---

Of the many paradoxes challenging assumptions about preferences I will survey two: the Sorites Paradox applied to preference transitivity and Allais’ paradox.

Sorites paradoxes are arguments that arise from the indeterminacy surrounding the limits of application of the predicates involved (for a general overview, see Hyde 2008). The Sorites scheme has been applied to RCT in order to cast doubt on the rationality of the transitivity of preference. Quinn's (1990) version goes as follows:

A person (call him *the self-torturer*) is strapped to a conveniently portable machine, which administers a continuous electric current. The device has 1,001 settings: 0 (off) and 1 … 1,000, of increasing current. The increments in current are so tiny that he cannot feel them. The self-torturer has time to experiment with the device so that he knows what each of the settings feels like. Then, at any time, he has two options: to stay put or to advance the dial one setting. However, he may advance only one step each week, and he may never retreat. At each advance he gets \$10,000.

Since the self-torturer cannot feel any difference in comfort between adjacent settings, he appears to have a clear and repeatable reason to increase the current each week. The trouble is that there are noticeable differences in comfort between settings that are sufficiently far apart. Eventually, he will reach settings that will be so painful that he would gladly relinquish his monetary rewards and return to zero.

The paradox lies in the conclusion that the self-torturer's preferences are intransitive. All things considered, he prefers 1–0, 2–1, 3–2, and so on, but certainly not 1,000–1. Furthermore, there seems to be nothing irrational about these preferences.

- The self-torturer's intransitive preferences seem perfectly natural and appropriate given his circumstances (Quinn 1990, p. 80).

If this were correct, the normative validity of the transitivity axiom would be in doubt. Defenders of transitivity argue that there is a mistake either in the conception of the decision situation or in the process of evaluation that leads to the intransitive preferences. Both Arntzenius and McCarthy (1997) and Voorhoeve and Binmore (2006) follow the first option in rejecting the implicit assumption that there is a “least-noticeable difference”: a magnitude of physical change so small that human beings always fail to detect a difference between situations in which a change smaller than this magnitude has or has not occurred.

Instead, they argue that it is rational for the self-torturer to take differences in long-run frequencies of pain reports into account. In other words, when repeatedly experimenting with the machine he may well experience different amounts of pain at the same notch. He will represent this information about how a notch feels by means of a distribution over different levels of pain: two notches “feel the same” only if they have the same distribution. Then it is implausible that all adjacent notches feel the same when the self-torturer runs through them in ascending order, and the intransitivity disappears.

Thus, Quinn's paradox has been treated as an apparent paradox: an implicit, illegitimate assumption of the derivation – of a “least-noticeable difference” – is exposed, and the dependency of the deductive conclusion on this assumption is shown.

*Allais' Paradox* (Allais 1953) sets up two specific choices between lotteries in order to challenge the sure-thing principle (an axiom in Savage's decision theory, related to the axiom of independence). This choice experiment is described in ► [Table 19.1](#). In this experiment, agents first choose between lotteries *A* and *B* and then between lotteries *C* and *D*.

RCT prescribes that agents choose *C* if they have chosen *A* (and vice versa), and that they choose *D* if they have chosen *B* (and vice versa). To see this, simply re-partition the prizes of the two problems as follows: Instead of “2,400 with certainty” in *B*, partition the outcome such that it reads “2,400 with probability 0.66” and “2,400 with probability 0.34.” Instead of “0 with

**Table 19.1**

Allais' two pairs of choices

Choice problem 1 – choose between:					
A:	\$2,500	With probability 0.33	B:	\$2,400	With certainty
	\$2,400	With probability 0.66			
	\$0	With probability 0.01			
Choice problem 2 – choose between:					
C:	\$2,500	With probability 0.33	D:	\$2,400	With probability 0.34
	\$0	With probability 0.67		\$0	With probability 0.66

**Table 19.2**

The re-described choice pairs

Choice problem 1* – choose between:					
A:	\$2,500	With probability 0.33	B*:	\$2,400	With probability 0.34
	\$2,400	With probability 0.66		\$2,400	With probability 0.66
	\$0	With probability 0.01			
Choice problem 2* – choose between:					
C*:	\$2,500	With probability 0.33	D:	\$2,400	With probability 0.34
	\$0	With probability 0.66		\$0	With probability 0.66
	\$0	With probability 0.01			

probability 0.67" in C, partition the outcome such that it reads "0 with probability 0.66" and "0 with probability 0.01." Of course, these are just redescriptions that do not change the nature of the choice problem. They are shown in [Table 19.2](#).

Through this redescription, we now have an outcome "2,400 with probability 0.66" in both A and B\*, and an outcome "0 with probability 0.66" in both C\* and D. According to the RCT independence condition, these identical outcomes can be disregarded in the deliberation, but once they are disregarded it becomes clear that option A is identical to option C\* and option B\* is identical to option D. Hence, anyone choosing A should also choose C and anyone choosing B should also choose D.

This result has been found both empirically and normatively challenging for RCT. (In sharp contrast to the RCT result, in an experiment involving 72 people, 82% of the sample chose B and 83% chose C (Kahneman and Tversky 1979).) On the normative level, many people seem to have intuitions contradicting the above conclusions:

- When the two situations were first presented, I immediately expressed preference for Gamble 1 [A] as opposed to Gamble 2 [B] and for Gamble 4 [D] as opposed to Gamble 3 [C], and I still feel an intuitive attraction to these preferences (Savage 1954, p. 103).

This empirical and normative challenge has given rise to a number of alternatives, including prospect theory (Kahneman and Tversky 1979), weighted utility (Chew 1983) and

rank-dependent expected utility (Quiggin 1982). However, the normative validity of these theories is often controversial. Many decision theorists have rather followed Savage, who despite his initial intuitions decided that the sure-thing axiom was, after all, correct. He arrived at this by redescribing Allais' choice situation in yet another form, observing a change of preference from  $C$  to  $D$  in this case, and concluded that "in revising my preferences between Gambles 3 [C] and 4 [D] I have corrected an error" (Savage 1954, p. 103). Decision theorists following Savage have thus treated Allais' paradox as a veridical paradox: the initial impression that the conclusion is absurd is explained away, and the theoretical conclusion is confirmed to be correct.

## Belief

---

I discuss only one paradox of belief here, namely the *Monty Hall problem*. It is posed as follows:

- ▶ Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice? (vos Savant 1990).

After picking a door at random, it may seem that it is rational to believe that the remaining door holds the car with probability  $\frac{1}{2}$ . After all, either the chosen door or the other one conceals the prize – so should both doors rather be assigned an equal probability of holding the car?

They should not. At first, before the contestant picks a door, it is rational for him or her to believe that the car is behind any of them with probability  $\frac{1}{3}$ , knowing that the host will be able to open a door not holding the prize since at least one of the other doors must conceal a goat. Therefore, when the host opens a door the contestant does not learn anything relevant to his or her belief in having chosen the winning door – it remains at  $\frac{1}{3}$ . Now the offered swap is equivalent to the opportunity of opening both other doors – and he or she should rationally believe that this offers a  $\frac{2}{3}$  probability of winning the car. Hence, it is advantageous to swap.

The Monty Hall problem clearly falls into the class of veridical paradoxes: the argument is correct and does not rely on implicit illegitimate assumptions, and the conclusion, despite appearances, is valid. Furthermore (unlike in most other paradoxes discussed here), the result can be experimentally confirmed: the frequency of the prize being behind the other door is observed indeed to converge to  $\frac{2}{3}$  (hence there is even a pragmatic confirmation of the normative intuition). The striking thing about this paradox is that the presentation of the correct answer initially created a huge outcry, not least from academically trained mathematicians and logicians (see <http://www.marilynvossavant.com/articles/gameshow.html> for a selection). The correct solution appeared to be false, but this appearance was explained away with the help of standard probability theory.

## Expected Utility

---

I will now discuss three paradoxes that challenge some fundamental assumptions about the rationality of expected utility maximization: the Ellsberg paradox, Newcomb's problem and the Envelope paradox.

The *Ellsberg paradox* (Ellsberg 1961) goes as follows. An urn contains 30 red balls and 60 other balls that are either black or yellow. You do not know how many black or yellow balls there are, but you do know that the total number of black balls plus the total number of yellow balls equals 60. The balls are well mixed so that each individual one is as likely to be drawn as any other. You are now given a choice between two gambles (☞ [Table 19.3](#)).

You are also given the choice between these two gambles with regard to a different draw from the same urn (☞ [Table 19.4](#)).

Standard (Bayesian) decision theory postulates that when choosing between these gambles, people assume a probability that the non-red balls are yellow versus black, and then compute the expected utility of the two gambles. This leads to the following line of reasoning. The prizes are exactly the same. Hence, according to expected utility theory, a rational agent (weakly) prefers *A* to *B* if and only if he or she believes that drawing a red ball is at least as likely as drawing a black ball. Similarly, a rational agent (weakly) prefers *C* to *D* if and only if he or she believes that drawing a red or yellow ball is at least as likely as drawing a black or yellow ball.

Now, if drawing a red ball is at least as likely as drawing a black ball, then drawing a red or yellow ball is also at least as likely as drawing a black or yellow ball. Thus, supposing that a rational agent (weakly) prefers *A* to *B*, it follows that he or she will also (weakly) prefer *C* to *D*, whereas supposing instead a weak preference for *D* over *C*, it follows that he or she will also weakly prefer *B* to *A*. When surveyed, however, most people (strictly) prefer Gamble *A* to Gamble *B* and Gamble *D* to Gamble *C*. Furthermore, they often insist on this choice, even if the theory is explained to them. Therefore, the normative validity of some assumptions of RCT seems in question.

Ellsberg's paradox poses an interesting challenge to RCT. On the one hand, some scholars have insisted that the standard solution is correct, making it a veridical paradox whose paradoxical impression is explained away. Fox and Tversky (1995), for example, offer an empirical explanation of why people may be biased in their decision-making through an impression of comparative ignorance. Such bias, of course, has no normative validity: it only explains why people have wrong intuitions about what should be chosen. On the other hand, others have argued that this ambiguity aversion is part of a rational decision, in a similar way as a risk aversion is (Schmeidler 1989). Such a position would suggest that the Ellsberg paradox is falsidical, brought about by assuming away the (normatively relevant) impact of ambiguity aversion.

■ **Table 19.3**  
Ellsberg's first pair of choices

Gamble A	Gamble B
You receive \$100 if you draw a red ball	You receive \$100 if you draw a black ball

■ **Table 19.4**  
Ellsberg's second pair of choices

Gamble C	Gamble D
You receive \$100 if you draw a red or yellow ball	You receive \$100 if you draw a black or yellow ball

*Newcomb's Problem* (Nozick 1969) involves an agent's choosing either an opaque box or the opaque and a transparent box. The transparent box contains one thousand dollars ( $\$T$ ) that the agent plainly sees. The opaque box contains either nothing ( $\$0$ ) or one million dollars ( $\$M$ ), depending on a prediction already made concerning the agent's choice. If the prediction was that the agent would take both boxes, then the opaque box will be empty, and if it was that the agent would take just the opaque box then the opaque box would contain a million dollars. The prediction is reliable. The agent knows all these features of the decision problem.

Table 19.5 displays the agent's choices and their outcomes. A row represents an option, a column a state of the world, and a cell an option outcome in a state of the world.

Standard RCT posits that a rational agent should choose the option that maximizes expected utility. This approach recommends taking only one box, for the following reasons. First, the prediction is reliable. In other words, if the agent chooses only one box, then the probability that "take one box" was predicted is high. Similarly, if the agent chooses two boxes, then the probability that "take two boxes" was predicted is high. Hence the probability of outcome  $\$M$  given that the agent chose only one box will be high, and the probability of outcome  $\$M + \$T$  given that the agent chose two boxes will be low – sufficiently low in most plausible cases for the expected utility of "taking one box" to be higher than that of "taking two boxes." Hence one-boxing is the rational choice according to the principle of expected-utility maximization.

Yet this recommendation violates two deeply entrenched intuitions. First, it violates the principle of dominance, according to which an agent prefers one action to another if he or she prefers every outcome of the first action to the corresponding outcomes of the second. The normative validity of dominance is widely agreed upon. Yet, two-boxing clearly dominates one-boxing in this sense. Consequently, RCT violates dominance.

Second, it violates the intuition that actions should be chosen on the basis of their causal effects rather than their probabilistic correlations to benefits or drawbacks. Because the prediction is made before the agent chooses, the choice has no causal impact on it, and the probabilistic correlation should not matter.

This analysis has motivated a reformulation of decision theory on causal rather than evidential grounds. In various accounts, causal decision theorists seek to represent causal influence with their probability functions rather than with mere probabilistic correlation (see e.g., Gibbard and Harper 1981; Skyrms 1980; Joyce 1999). They clearly see Newcomb's problem as a falsidical paradox based on the misspecification of a decision maker's relevant beliefs.

Some authors opposing the causal approach hold that it yields the wrong choice in Newcomb's problem, in other words two-boxing rather than one-boxing. Horgan (1985)

Table 19.5  
Newcomb's problem

	Prediction of one-boxing	Prediction of two-boxing
Take one box	$\$M$	$\$0$
Take two boxes	$\$M + \$T$	$\$T$

and Horwich (1987), for example, argue that one-boxers fare better than two-boxers, and that one-boxing is therefore the rational choice of action. They both see Newcomb's problem as a veridical paradox, and propose explaining away the conflicting intuition about dominance and causal influence with reference to pragmatic success. Causal decision theorists, in turn, reject the relevance of these pragmatic considerations for the validity of the above intuitions. They insist that Newcomb's problem is an unusual case, which rewards irrationality: one-boxing is irrational even if one-boxers prosper.

The *two-envelope paradox* goes as follows. You are asked to make a choice between two envelopes. You know that one of them contains twice the amount of money as the other, but you do not know which one. You arbitrarily choose one envelope – call it Envelope A – but do not open it. Call the amount of money in that envelope  $X$ . Since your choice was arbitrary, there is a 50–50 chance that the other envelope (Envelope B) will contain more money, and a 50–50 chance that it will contain less. Would you now wish to switch envelopes?

Calculating the apparent expected value of switching proceeds as follows. Switching to  $B$  will give you a 50% chance of doubling your money and a 50% chance of halving it. Thus it seems that the expected value of switching to  $B$  is  $E(Y-X) = 0.5*1/2X + 0.5*2X - X = 0.25X$ . Hence, switching to  $B$  will give you a 25% higher expected return than sticking with  $A$ . This seems absurd. First, many people have an intuition that one should be indifferent between  $A$  and  $B$  as long as the envelope remains unopened. Second, once you have switched to  $B$  in line with the above argument, a symmetrical calculation could persuade you to switch back to  $A$ . Therein lies the paradox.

It is now widely agreed that the expected gain from switching,  $E(Y-X)$ , is mathematically undefined because the value of the infinite sum of all probability-weighted values of  $Y-X$  depends on the order of summation (Meacham and Weisberg 2003). However, the conclusions from this observation differ widely. Clark and Shackel (2000) argue that there is a “correct” order of summation, which results in a zero infinite sum, and that this result justifies indifference before opening the envelope. In contrast, Meacham and Weisberg (2003) express reservations about selecting the “correct” order of summation: because the expected gain from switching is undefined, standard decision theory does not rank switching against keeping.

Dietrich and List (2005) go along a different route and offer an axiomatic justification for indifference before opening without appeal to infinite expectations. They supplement standard decision theory with an additional axiom, the “indifference principle,” according to which if two lotteries have identical distributions, a rational agent is indifferent between them. From this they are able to deduce a justification for indifference before opening. All three of these responses, although formulated against each other, consider the two-envelope paradox falsidical: Clark and Shackel and Dietrich and List introduce additional assumptions, which yield the intuitive conclusion, whereas Mechaem and Weisberg insist that the argument is fallacious, and no conclusion is warranted from the given assumptions.

## Strategic Interaction

Game theory is closely related to RCT. Although it requires certain assumptions beyond those of the standard RCT axioms, its models give additional significance to those standard RCT axioms that also play a role in game theory. For this reason, I include two game-theoretic paradoxes here: the Prisoners' dilemma and the paradox of common knowledge.

The One-Shot *Prisoners' Dilemma* has attracted much attention because standard game-theoretic solution concepts unanimously advise each player to choose a strategy that will result in a Pareto-dominated outcome. It goes as follows.

Two gangsters who have been arrested for robbery are placed in separate cells. Both care much more about their personal freedom than about the welfare of their accomplice. A prosecutor offers each the following deal: choose to confess or remain silent. If one prisoner confesses and the accomplice remains silent all charges against the former are dropped (resulting in a utility of 3 in [Table 19.6](#) – the first number in each cell is the utility of this outcome for the player who chooses between rows, and the second number, the utility for the player who chooses between columns). If the accomplice confesses and the first prisoner remains silent however, the latter will do time (utility 0 in [Table 19.6](#)). If both confess, each will get reduced sentences (utility 1), and if both remain silent the prosecutor has to settle for token sentences for firearms possession (utility 2) (for extensive discussion and a literature review, see Kuhn 2009).

The choice situation is solved by appeal to a simple dominance argument. For each player, if the other player stays silent it is better to confess than to stay silent. If the other player confesses, it is also better to confess than to stay silent. Hence, no matter what the other player does, it is always better to confess.

This result is often described as paradoxical in the following sense. The outcome obtained when both confess, although it is rational for each to do so, is worse for each than the outcome they would have obtained had both remained silent. Both would prefer to reach the outcome "stay silent, stay silent," but their individually rational actions led them to the inferior result "confess, confess." (To add some more urgency to this example, consider the structurally similar problem of the "tragedy of the commons," according to which multiple individuals acting independently and rationally will ultimately deplete a shared limited resource even when it is clear that it is not in anyone's long-term interest for this to happen (Hardin 1968).) How can such an inferior outcome be the result of rational decisions?

Some authors argue that the Prisoners' Dilemma indeed exposes a limitation of RCT rationality. Gauthier (1986), for example, suggests that, instead of always confessing, it would be rational for players to commit to playing cooperatively when faced with other cooperators who are equally committed to not exploiting one another's good will. This argument crucially depends on player confidence in that most players are clearly identifiable as being committed to cooperating or not. Whether such a belief can be rationally justified is questionable, and with it the whole solution to the dilemma.

The majority of authors see no conceptual problem in the Prisoners' Dilemma. The assumptions say nothing about the necessary selfishness of the players (charity organizations may also find themselves in such situations!), and no other illegitimate assumptions are

[\*\*Table 19.6\*\*](#)

**The Prisoners' dilemma**

	Stay silent	Confess
Stay silent	2,2	0,3
Confess	3,0	1,1

evident. The argument itself is valid, and the conclusion is not contradicted by normative intuitions. The only problem is terminological: some people chafe against the idea that the conclusion is supposed to be the result of rational decision-making. However, they simply subscribe to a different concept of rationality that RCT does not support. Hence the Prisoners' Dilemma is a veridical paradox, whose paradoxical nature relies on terminological ambiguity.

*Backward induction* is the process of reasoning backward in time, from the end of a problem, to determine a sequence of optimal actions. It proceeds by first considering the latest time a decision can be made and choosing what to do in any situation at that time. One can then use this information to determine what to do at the second-to-last time for the decision. This process continues backward until one has determined the best action for every possible situation at any point in time.

Let us take a concrete example, the “centipede game.” This game progresses from left to right in Fig. 19.1. Player 1 (female) starts at the extreme left node, choosing to end the game by playing *down*, or to continue (giving player 2, male, the choice) by playing *right*. The payoffs are such that at each node it is best for the player whose move it is to stop the game if and only if he or she expects it to end at the next stage if he or she continues (if the other player stops the game or if it is terminated). The two zigzags stand for the continuation of the payoffs along those lines. Now backward induction advises resolving the game by starting at the last node *z*, asking what player 2 would have done had he ended up there. A comparison of player 2's payoffs for the two choices implies that he should have rationally chosen *down*. Given common knowledge of rationality, the payoffs that result from this choice of *down* can be substituted for node *z*. Let us now move backwards to player 1's decision node. What would she have done had she ended up at node *y*? Given player 2's choice of *down*, she would have chosen *down*, too. This line of argument then continues all the way back to the first node. Backward induction thus recommends player 1 to play *down* at the first node.

What, then, should player 2 do if he actually found himself at node *x*? Backward induction tells him to play “*down*,” but backward induction also tells him that if player 1 were rational he would not be facing the choice at node *x* in the first place. This is not a problem in that Backward induction predicts that player 2 will never find himself at *x* unless both players are irrational.

Yet what does this imply for the Backward-induction reasoning process itself? Backward induction requires the players to *counterfactually* consider out-of-equilibrium play. For example, player 1, according to Backward induction, should choose *down* at node 1, because she knows that player 2 would have chosen *down* at node 2, which in turn she knows because she would have chosen *down* at node 3, and so on, because ultimately she knows that player 2 would have chosen *down* at *z*. She knows this because she knows that player 2 is rational,

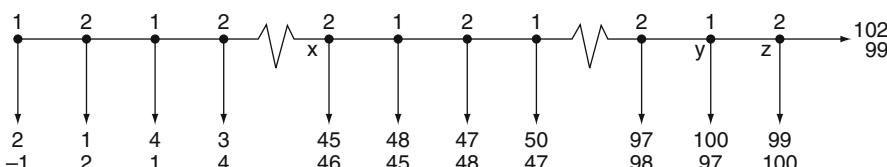


Fig. 19.1  
The centipede game

and that player 2 knows that she is rational, and so on. In the language of game theory, because rationality is *common knowledge* amongst the players, backward induction applies (for more on common knowledge, including a formal treatment of this paradox, see Vanderschraaf and Sillari 2009).

Given common knowledge of rationality, each player can affirm the counterfactual “A rational player finding himself or herself at any node in the centipede would choose *down*.” Yet we also concluded that if a player *finds* himself or herself at any node with an index number larger than two then both player and opponent know that they *are* not rational. What if that conclusion also made true the counterfactual “if a player *found* himself or herself at any node with an index larger than two then both player and opponent *would* know that they are not rational”? If it did, Backward induction would break down: it requires (1) common knowledge and (2) counterfactual consideration of how players would choose if they found themselves at nodes with indices larger than two. However, (2) implies that both player and opponent would know that they are not rational, contradicting (1). Herein lies the paradox (Pettit and Sugden 1989; Bicchieri 1989).

This is an intensely discussed problem in game theory and philosophy. There is space here only to sketch two possible solutions. According to the first, common knowledge of rationality implies backward induction in games of perfect information (Aumann 1995). This position is correct in that it denies the connection between the indicative and the counterfactual conditional. Players have common knowledge of rationality, and they are not going to lose it regardless of the counterfactual considerations they engage in. Only if common knowledge were not immune to evidence, and would be revised in the light of the opponents’ moves, might this sufficient condition for backward induction run into the *conceptual problem* sketched above. However, common knowledge, by definition, is not revisable, and thus the argument has to assume a *common belief* in rationality instead. If one looks more closely at the versions of the above argument (e.g., Pettit and Sugden 1989) it becomes clear that they employ the notion of common belief rather than common knowledge. Hence, the backward-induction paradox is only apparent: the argument that led to the seemingly contradictory conclusion is unsound.

The second potential solution obtains when one shows, as Bicchieri (1993, Chap. 4) does, that limited knowledge (and *not* common knowledge) of rationality and of the structure of the game suffice for backward induction. All that is needed is that a player at each information set knows what the next player to move knows. This condition does not get entangled in internal inconsistency, and backward induction is justifiable without conceptual problems. In that case, the backward-induction paradox is falsidical.

## Further Research

I have surveyed eight paradoxes of RCT. There is considerable divergence among them, under a rather rough classificatory scheme, even in this small selection. First, there are the veridical paradoxes, like the Prisoner’s dilemma, the paradoxical nature of which rests merely on terminological ambiguity. Ways of explaining away the paradoxical appearance of other veridical paradoxes such as the Monty Hall problem are obvious, but are baffling to the novice. They can still serve an educational purpose, however, in that studying them clarifies the meaning of the assumptions and the derivation of the conclusion.

Second, there are (relatively) clear cases of falsidical paradoxes, such as the two-envelope paradox. Here, there is a clear research result: RCT needs revision.

Third, there are some clear cases of apparent paradoxes, such as the self-torturer, in which the whole bluster is caused by a fallaciously set up argument.

Finally, there are cases on which researchers cannot agree. These include Newcomb's problem, Allais' and Ellsberg's paradox, which vacillates between veridical and falsidical assessment, and the Backward-induction paradox, which vacillates between falsidical and apparent assessment. In all these cases the verdict is still open as to whether they necessitate RCT revision or not. Hence their continuing examination is part of active research in this area.

## References

---

- Allais M (1953) Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Américaine. *Econometrica* 21:503–546
- Arntzenius F, McCarthy D (1997) Self torture and group beneficence. *Erkenntnis* 47(1):129–144
- Aumann R (1995) Backward induction and common knowledge of rationality. *Game Econ Behav* 8:6–19
- Bicchieri C (1989) Self refuting theories of strategic interaction: a paradox of common knowledge. *Erkenntnis* 30:69–85
- Bicchieri C (1993) Rationality and coordination. Cambridge University Press, Cambridge
- Chew SH (1983) A generalization of the quasilinear mean with application to the measurement of income inequality and decision theory resolving the Allais paradox. *Econometrica* 51:1065–1092
- Clark M, Shackel N (2000) The two-envelope paradox. *Mind* 109(435):415–442
- Daniels N (2008) Reflective equilibrium. In: Zalta EN (ed) The Stanford encyclopedia of philosophy, fallth edn. The Metaphysics Research Lab, Stanford, Available online: <http://plato.stanford.edu/archives/fall2008/entries/reflective-equilibrium/>
- Diekmann A, Mitter P (1986) Paradoxical effects of social behavior: essays in honor of Anatol Rapoport. Physica-Verlag, Heidelberg and Vienna, Available online: <http://www.socio.ethz.ch/vlib/pepb/index>
- Dietrich F, List C (2005) The two-envelope paradox: an axiomatic approach. *Mind* 114:239–248
- Ellsberg D (1961) Risk, ambiguity, and the savage axioms. *Q J Econ* 75(4):643–669
- Fox CR, Tversky A (1995) Ambiguity aversion and comparative ignorance. *Q J Econ* 110(3):585–603
- Gauthier D (1986) Morals by agreement. Oxford University Press, Oxford
- Gibbard A, Harper W (1981) Counterfactuals and two kinds of expected utility. In: Harper W, Stalnaker R, Pearce G (eds) Ifs: conditionals, belief, decision, chance, and time. Reidel, Dordrecht, pp 153–190
- Guala F (2000) The logic of normative falsification: rationality and experiments in decision theory. *J Econ Methodol* 7(1):59–93
- Hansson SO, Grüne-Yanoff T (2009) Preferences. In: Zalta EN (ed) The stanford encyclopedia of philosophy, springth edn. The Metaphysics Research Lab, Stanford, Available online: <http://plato.stanford.edu/archives/preferences/>
- Hardin G (1968) The tragedy of the commons. *Science* 162(3859):1243–1248
- Hargreaves-Heap S, Hollis M, Lyons B, Sugden R, Weale A (1992) The theory of choice: a critical guide. Blackwell, Oxford
- Horgan T (1985) Counterfactuals and Newcomb's problem. In: Campbell R, Sowden L (eds) Paradoxes of rationality and cooperation: Prisoner's dilemma and Newcomb's problem. University of British Columbia Press, Vancouver, pp 159–182
- Horwich P (1987) Asymmetries in time. MIT Press, Cambridge, MA
- Hyde D (2008) Sorites paradox. In: Zalta EN (ed) The stanford encyclopedia of philosophy, fallth edn. The Metaphysics Research Lab, Stanford, Available online: <http://plato.stanford.edu/archives/fall2008/entries/sorites-paradox/>
- Jeffrey R (1990) The logic of decision, 2nd edn. University of Chicago Press, Chicago
- Joyce J (1999) The foundations of causal decision theory. Cambridge University Press, Cambridge
- Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. *Econometrica* 47:263–291
- Koops R (1992) Paradoxes of belief and strategic rationality. Cambridge University Press, Cambridge
- Kuhn S (2009) Prisoner's dilemma. In: Zalta EN (ed) The Stanford encyclopedia of philosophy, springth edn.

- The Metaphysics Research Lab, Stanford, Available online: <http://plato.stanford.edu/archives/spr2009/entries/prisoner-dilemma/>
- Luce RD, Raiffa H (1957) Games and decisions: introduction and critical survey. Wiley, New York
- Lycan WG (2010) What, exactly, is a paradox? *Analysis* 70(4):615–622
- Mas-Colell A, Whinston MD, Green JR (1995) Microeconomic theory. Oxford University Press, New York
- Meacham CJG, Weisberg J (2003) Clark and Shackel on the two-envelope paradox. *Mind* 112(448):685–689
- Nozick R (1969) Newcomb's problem and two principles of choice. In: Rescher N (ed) Essays in honor of Carl G. Hempel. Reidel, Dordrecht, pp 114–146
- Pettit P, Sugden R (1989) The backward induction paradox. *J Philos* 86:169–182
- Quiggin J (1982) A theory of anticipated utility. *J Econ Behav Organ* 3(4):323–343
- Quine WVO (1966) The ways of paradox. In: The ways of paradox and other essays. Random House, New York, pp 3–20, Original published as Paradox (1962) in *Scientific American* 206(4):84–95
- Quinn WS (1990) The puzzle of the self-torturer. *Philos Stud* 59(1):79–90
- Resnik MD (1987) Choices: an introduction to decision theory. University of Minnesota Press, Minneapolis
- Richmond C, Sowden L (eds) (1985) Paradoxes of rationality and cooperation: Prisoners' dilemma and Newcomb's problem. University of British Columbia Press, Vancouver
- Sainsbury RM (1988) Paradoxes. Cambridge University Press, Cambridge
- Savage LJ (1954) The foundations of statistics. Wiley, New York
- Schmeidler D (1989) Subjective probability and expected utility without additivity. *Econometrica* 57:571–587
- Skyrms B (1980) Causal necessity: a pragmatic investigation of the necessity of laws. Yale University Press, New Haven
- Vanderschraaf P, Sillari G (2009) Common knowledge. In: Zalta EN (ed) The stanford encyclopedia of philosophy, strength edn. The Metaphysics Research Lab, Stanford, Available online: <http://plato.stanford.edu/archives/spr2009/entries/common-knowledge/>
- von Neumann J, Morgenstern O (1947) The theory of games and economic behavior, 2nd edn. Princeton University Press, Princeton
- Voorhoeve A, Binmore K (2006) Transitivity, the Sorites paradox, and similarity-based decision-making. *Erkenntnis* 64(1):101–114
- vos Savant M (1990) Ask marilyn column. *Parade magazine*, 9 Sept 1990, p 16

# 20 Multi-Attribute Approaches to Risk

*Paul Weirich*

University of Missouri, Columbia, MO, USA

<b>Introduction .....</b>	<b>518</b>
<b>History .....</b>	<b>519</b>
Multi-Attribute Utility Analysis .....	519
Related Methods .....	521
<b>Current Research .....</b>	<b>523</b>
Government Regulation .....	523
Risk as an Attribute .....	524
Risk-Return Assessment of Investments .....	527
Multi-Attribute Utility Functions .....	528
Intrinsic-Utility Analysis .....	532
Trustee Decisions .....	534
<b>Further Research .....</b>	<b>536</b>
Aggregation Functions .....	536
Realism .....	537
Incomparable Attributes .....	539
Outlook .....	542

**Abstract:** A person who adopts a risky option undertakes a risk. However the option turns out, the person experiences a risk. A typical person is averse to that risk, and so, if the option's adoption is rational, the option has attributes that compensate for the risk it involves. A multi-attribute assessment of a risky option examines the option's attributes, scores the option with respect to them, and checks whether its overall score justifies the option despite the risk it involves.

The multi-attribute approach to risk is an elaboration of Bayesian decision theory. It assumes utility maximization as a standard of rationality and offers a method of assessing an option's utility. Given uncertainty about an option's outcome, it assumes that the option's utility equals the option's expected utility. To an expected-utility analysis of an option's utility using the probabilities and utilities of the option's possible outcomes, it adds an analysis of every possible outcome's utility. It breaks down a possible outcome's utility into the possible outcome's utility with respect to various attributes.

Some versions of the multi-attribute approach to risk treat risk as an attribute, and other versions fold an evaluation of an option's risk into its evaluation with respect to other attributes. This chapter presents the multi-attribute approach to risk, its most prominent versions, objections to the approach, and ways of addressing the objections.

Although descriptive accounts of people's decisions may use a multi-attribute analysis of an option's utility, this chapter examines only evaluative forms of multi-attribute analyses of utility. The evaluative forms use multiple attributes to evaluate the options in a decision problem. The literature also calls these forms of analysis normative or prescriptive.

In a decision problem, the multi-attribute approach to risk provides a transparent evaluation of options and a transparent evaluation of a decision among the options. The approach's transparency makes it a valuable tool for decisions that require a public justification, such as decisions that trustees make for their clients, in particular, decisions that government regulatory agencies make on behalf of the public. To display the usefulness of the multi-attribute approach to risk, this chapter reviews its application to decisions that a trustee makes for a client.

## Introduction

---

Each spring, graduating seniors bent on academic careers receive offers of admission into graduate programs. A student with several offers evaluates each with respect to features such as the quality of the program, the amount of financial support, and the geographic location of the program. The best program may offer the least financial support and have the worst location. Is it sensible to pick the second best program if it offers the best financial support and has the best location? Decisions require trade-offs between goals such as entering a top program, being fully funded, and living in an attractive area. The multi-attribute approach to risk is a method of making trade-offs, including trade-offs between risk and other factors.

An agent in a decision problem has a set of options and resolves the problem by selecting an option. The multi-attribute approach to decisions, in its canonical form, evaluates each option with respect to multiple attributes and uses these evaluations to reach an overall evaluation of the option. Then it uses options' overall evaluations to reach a decision. When some options in a decision problem involve risks, the multi-attribute approach may use risk as an attribute. A very risky option may receive a low mark with respect to that attribute. Its overall evaluation is low unless better marks with respect to other attributes compensate for its low mark.

concerning risk. The multi-attribute approach to decision problems with risky options constitutes the multi-attribute approach to risk. Some government agencies take this approach to regulatory decisions, which often involve trade-offs between safety and other social values. This chapter explains the multi-attribute approach to risk, illustrates its value for decisions that trustees make for clients, and explores ways of improving the approach to risk.

The chapter adopts a broadly Bayesian method of making decisions. It assumes that an agent assigns utilities to options and, if rational, adopts an option of maximum utility among options. Moreover, when an option's outcome is uncertain, it assumes that the agent assigns a probability and utility to each of the option's possible outcomes. If the agent is rational, the utility she assigns to an option is a probability-weighted average of the utilities she assigns to its possible outcomes. See in this handbook, [Chap. 15, A Rational Approach to Risk? Bayesian Decision Theory](#).

Bayesian decision principles govern ideal agents, who have no cognitive limits and know all logical and mathematical truths. To simplify exposition of the multi-attribute approach to risk, this chapter focuses on rational ideal agents. For such agents, an option's utility equals its expected utility. Because for the agents the chapter treats, this equality holds, the chapter generally does not distinguish an option's utility and its expected utility. It also assumes, as is common in Bayesian decision theory, that an option's expected utility has the same value with respect to every partition of the option's possible outcomes.

Although in their standard forms Bayesian decision theory and the multi-attribute approach to risk treat an ideal agent in an ideal decision problem, they are useful guides for a real agent in a real decision problem. To show how to make precise their extensions to realistic cases, section [Further Research](#) briefly sketches ways of relaxing standard idealizations and formulating general decision principles that cover all cases.

## History

---

The multi-attribute approach to decisions about risk uses a traditional method of evaluating an option, namely, the method of considering the pros and cons of the option's realization, weighing the pros and cons, and finally adding the pros and cons to obtain the option's comprehensive evaluation. The multi-attribute approach uses an option's scores with respect to attributes as pros and cons, weights attributes according to importance, and adds an option's weighted scores with respect to attributes to obtain the option's comprehensive evaluation. The approach gains precision by being explicit about the set of attributes, their weights, and the method of combining an option's evaluation with respect to the attributes into an overall evaluation of the option.

## Multi-Attribute Utility Analysis

---

In a decision problem, such as selecting dinner from a menu, an agent has a set of options. The agent resolves the problem by adopting an option from the set. An option is an act that the agent fully controls. If an agent directly controls an option, then rationality approves of the option only if it compares well with other options that the agent also directly controls. Rationality requires an option at the top of an agent's preference ranking of options.

In the decision problems this section considers, comparisons of options depend on how the options score with respect to a set of attributes. The main components of a multi-attribute evaluation of an option are: (1) a set of attributes with respect to which the option is evaluated and (2) a method of combining evaluations with respect to attributes to obtain an option's overall evaluation. Options' overall evaluations yield a ranking of the options, and the ranking identifies rational options. They are the options at the top of the ranking.

Numbers called utilities represent options' evaluations. An assignment of a utility to an option indicates how it compares with other options. This chapter takes utility as strength of desire held all things considered and takes utility as attaching to a proposition because desire attaches to a proposition. A rational ideal agent's degrees of desire comply with the principles of utility theory. Rationality gives them the usual structure of utilities.

According to another common interpretation of utilities, they are assignments of numbers to options and to outcomes of options that obey the expected utility principle and that represent preferences among options. Representation theorems identify conditions on preferences that suffice for preferences having such a representation and its being unique given a choice of scale for utilities. This interpretation makes the expected utility principle hold as a matter of definition. To make the expected utility principle an interesting normative requirement instead of a definitional truth, this chapter takes utilities as degrees of desire rather than as artifacts of a mathematical representation of preferences.

Given that rational degrees of desire satisfy the expected utility principle, the methods of proving the representation theorems show how to measure utilities on a utility scale. A rational ideal agent's preferences among options involving chance, because of the structure of these preferences, permit an outside evaluator to infer the agent's utility assignments to options and to their possible outcomes. In some cases, comparisons among options are so richly structured that, given the utility scale, only one utility representation satisfies the expected utility principle. In this case, it yields the agent's utility assignment. Although in special cases an agent's utility assignments are inferable from her preferences, in the cases this chapter treats, her utility assignments to possible outcomes generate her utility assignments to options and her preferences among options. Her utility assignments to an option's possible outcomes explain her utility assignment to the option, and her utility assignments to options explain her preferences among options.

A multi-attribute evaluation of an option generates the option's utility from its utilities with respect to a set of attributes. It assigns a utility to the option with respect to each attribute and then combines those restricted utilities to obtain the option's comprehensive utility. Because the procedure uses utilities to indicate comparisons of options, theorists call the procedure multi-attribute utility analysis and call its rationale multi-attribute utility theory. Keeney and Raiffa ([1976]1993) present a canonical version of multi-attribute utility analysis, with a thorough explanation of its steps and with many illustrations. Mullen and Roth (1991, Chaps. 3, 7); Kleindorfer et al. (1993, Sect. 4.2.3, Appendix C), and Ellis (2006) present valuable summaries of multi-attribute utility theory.

A utility scale used to represent an option's evaluation with respect to an attribute may be either ordinal or cardinal. If the utility scales for all attributes are cardinal scales, an aggregation of utilities with respect to attributes may be additive and may weight utilities with respect to attributes before adding them.

To illustrate the procedure of evaluation, consider an evaluation of travel by train and travel by plane. A multi-attribute evaluation of these modes of transportation may use the attributes of speed and comfort. A cardinal scale fits speed. The plane is three times as fast as the train, say.

The scale may use the utility of the train's speed as the unit for utility with respect to speed. Then on the scale for speed, taking the train has a utility of 1, and taking the plane has a utility of 3. An ordinal scale fits comfort. Although the train is more comfortable than the plane, it is typically not some number of times as comfortable. Suppose, however, that for an especially sensitive passenger, the train is twice as comfortable as the plane. Using the utility of comfort on the plane as the unit for the scale of utility with respect to comfort, the utility of taking the plane is 1, and the utility of taking the train is 2. In this special case, a cardinal scale fits both attributes.

Imagine that for the passenger, speed is four times as important as comfort. Using the weight of comfort as a unit for the scale for weights, the weight of comfort is 1 and the weight of speed is 4. This quantitative weighting of attributes is a stretch, but simplifies the illustration. Section  [Further Research](#) considers how to relax such simplifying assumptions.

 [Table 20.1](#) displays a multi-attribute assessment of the two modes of travel. It uses the weights of attributes to combine an option's utilities with respect to attributes into a comprehensive utility. Because taking the plane has the greater comprehensive utility, the decision principle to maximize utility recommends it.

A multi-attribute utility analysis often takes attributes of options to be types that various options realize in different ways. Variables represent attribute types that many options may instantiate. A value of the variable represents an option's instantiation of the attribute. For example, suppose that in a decision problem the options are jobs. An attribute of the options is income, and a variable represents income for a job. The various jobs available differ in income. A job's income supplies a realization of the attribute, that is, a value for the income variable. Deliberation may compare options according to their instantiations of an attribute type; for example, it may compare jobs according to the incomes the jobs provide.

Although taking attributes as types is common, attributes may be tokens, or instances of properties. A multi-attribute analysis may take a job's particular income as an attribute without treating its income as an instance of an attribute-type. Then the set of attributes for one option may differ from the set of attributes for another option. Although using the same set of attribute types for all options facilitates comparisons of options, an option's evaluation may use attributes that pertain to it exclusively. For each option, an evaluation may use a different set of attributes. The attributes for an option may be the option's particular features.

## Related Methods

---

The multi-attribute approach to risk may, for simplicity, compare options pair-wise without first generating an overall evaluation of each option. For example, one option may surpass all other options with respect to every attribute. In that case, the option is atop any ranking of options. Its comparison with other options need not evaluate each option with respect to all attributes and

 **Table 20.1**

**Multi-attribute assessment of modes of travel**

	Utility for speed	Utility for comfort	Weighted utility
Train	1	2	$4(1) + 1(2) = 6$
Plane	3	1	$4(3) + 1(1) = 13$

thereby generate an overall score for the option. Reaching a comparison of two options does not require assigning scores to each option. However, this chapter, for simplicity, considers only decision problems in which an option's evaluations with respect to attributes generate the option's overall evaluation, and the options' overall evaluations generate comparisons of options.

The main alternative to a multi-attribute evaluation of the options in a decision problem is a single-attribute evaluation of the options. A single-attribute evaluation uses utility as an attribute and assigns a utility to each option. Its evaluation of an option does not break down the option's utility into its utilities with respect to multiple attributes. Options' utilities generate comparisons of options. An option of maximal utility tops the ranking of options and constitutes a rational choice.

Although a single-attribute evaluation of options is simpler, a multi-attribute evaluation better justifies a decision. A single-attribute evaluation of an option does not offer an account of the option's utility, unless it derives the option's utility using an expected-utility analysis, and then does not offer an account of the utilities of the option's possible outcomes. A multi-attribute evaluation of an option shows how the option's utility, or a possible outcome's utility, arises from its utilities with respect to the option's or possible outcome's attributes and their weights in the option's or possible outcome's overall assessment.

This chapter, unlike some presentations of multi-attribute utility theory, distinguishes multi-attribute utility analysis from multidimensional utility analysis. Multi-attribute utility analysis divides an option's comprehensive utility into the utilities the option has with respect to a set of attributes. Its division uses the attributes as a dimension of analysis. The attributes form just one dimension of analysis, however. Other forms of utility analysis use other dimensions of analysis. For example, expected-utility analysis divides an option's utility into the utilities of chances for the option's possible outcomes. The possible outcomes form another dimension of utility analysis. Also, utilitarianism divides an option's utility into the utilities the option generates for the individuals it affects. The individuals form a dimension of utility analysis. A multidimensional analysis of an option's utility uses multiple dimensions of analysis, for example, attributes, possible outcomes, and people. Weirich (2001) offers an account of multidimensional utility analysis.

A multi-attribute analysis of an option's utility uses the dimension of attributes. It is appropriate when the option is sure to have a certain outcome, and the option will affect only its agent. Given uncertainty, a multi-attribute utility analysis evaluates each of an option's possible outcomes according to a set of attributes. Then it obtains the option's utility as a probability-weighted sum of the utilities of its possible outcomes. Because the last step constitutes a traditional expected-utility analysis, the process is also a multidimensional form of utility analysis. The literature counts as a multi-attribute utility analysis any analysis of an option's utility that uses the dimension of attributes, possibly along with other dimensions such as the dimension of possible outcomes.

This chapter generally treats multi-attribute utility analysis in its pure form, using only the dimension of attributes, and only occasionally treats its combination with expected-utility analysis. When it entertains an expected-utility analysis of an option's utility using probabilities of the option's possible outcomes, it adopts the standard Bayesian interpretation of probabilities, according to which probabilities are rational degrees of belief and attach to propositions as belief do. It uses propositions to represent possible outcomes. In some cases a rational ideal

agent's subjective probabilities may be inferred from the agent's preferences among gambles, but nonetheless the subjective probabilities are causes of the preferences in the cases this chapter treats.

## Current Research

---

The previous section introduced multi-attribute utility theory. This section reviews the theory's application to risk, a topic of current research. A major issue is whether to treat risk as an attribute or to assume that utility assignments with respect to other attributes handle risk.

### Government Regulation

---

An option's assessment with respect to attributes is the cornerstone of the multi-attribute approach to risk. As an application of the approach, consider government regulation to reduce risk. Regulatory goals provide attributes for assessing regulations. A government agency's decision to regulate a risky technology typically involves trade-offs between regulatory goals, such as safety and freedom. For example, the Food and Drug Administration's decision to halt genetic therapy for severe combined immunodeficiency, or "Bubble Boy" disease, because of cases in which the therapy caused leukemia, promotes safety but limits freedom by denying a treatment option. Because a regulation makes trade-offs among goals, its justification assesses the regulation with respect to each relevant goal and then combines the assessments. The regulation's assessments with respect to regulatory goals and the goals' relative weights yield the regulation's overall assessment. The regulation's comparisons with other options, according to overall assessments of options, settle whether, among options that an agency's mandate allows, the agency's regulation is optimal.

The set of attributes with respect to which the multi-attribute approach to risk assesses an option may use specific objectives in place of general goals and so may include several types of benefit and several types of risk. For example, a new ointment may create multiple benefits and multiple risks because it may soften skin and remove wrinkles but also in rare cases damage capillaries and weaken defenses against skin cancer. A multi-attribute evaluation of marketing the ointment may, as attributes, use soft skin, fewer wrinkles, risk of capillary damage, and risk of skin cancer. It may assess a regulatory agency's options with respect to each of these attributes and then combine the options' relative assessments into the options' overall assessments. The agency's decision to authorize marketing the new ointment then uses a multi-attribute assessment with respect to specific objectives rather than with respect to general regulatory goals.

The grain of the set of attributes is a matter of practicality. Although a general goal is risk reduction, an evaluation may achieve accuracy more easily by focusing on specific objectives concerning reduction of various types of risk. An apt choice of a set of attributes facilitates a multi-attribute assessment of options.

Some features of the set of attributes are critical for a reliable application of the multi-attribute approach. Keeney and Raiffa ([1976]1993, p. 51) offer this advice about selection of attributes: "It is important in any decision problem that the set of attributes be *complete*, so that

it covers all the important aspects of the problem; *operational*, so that it can be meaningfully used in the analysis; *decomposable*, so that aspects of the evaluation process can be simplified by breaking it down into parts; *nonredundant*, so that double counting of impacts can be avoided; and *minimal*, so that the problem dimension is kept as small as possible.” They acknowledge (p. 53) that for a specific decision problem a suitable set of attributes is not unique, and they suggest choosing a set of attributes that serves the purposes of an analysis of the decision problem and that facilitates assessments of probabilities and utilities of options’ possible outcomes. Selection of a set of attributes is partly a pragmatic matter.

Keeney and Raiffa’s second criterion for a set of attributes requires them to be operational. The term operational has different meanings in different disciplines. By operational attributes, Keeney and Raiffa mean attributes that can be measured. They do not mean attributes that have operational definitions in the philosophical sense. Hardly any interesting attributes have precise, accurate definitions that reduce the attributes to the results of observationally verifiable operations.

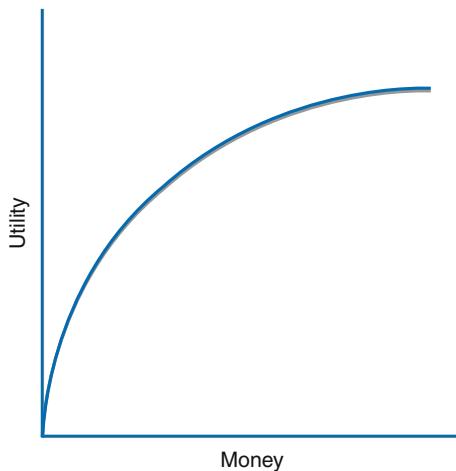
Peterson (2007, pp. 74–75) notes that some methods of aggregating attributes yield different rankings of options depending on which set of attributes they use. For instance, a ranking of treatments for a disease may change if in a population a possible side effect, cardiotoxicity, is divided into cardiotoxicity for each of two subpopulations. The side effect may emerge with significant frequency only in one of the subpopulations. Such a change in ranking, if the different sets of attributes meet all relevant criteria, makes the multi-attribute approach yield ambiguous directions. A thorough defense of multi-attribute assessments of options must show that all properly selected sets of attributes yield the same ranking of options.

The multi-attribute approach to regulation of risk has the following main steps. Given the emergence of a potentially beneficial but risky technology, it evaluates each regulatory option, including not imposing a regulation, with respect to a set of regulatory goals, and then combines those evaluations to obtain the option’s overall evaluation. It uses options’ overall evaluations to compare the options and to select one. It explicitly evaluates each option with respect to the goal of risk reduction, or it considers each option’s risk during its evaluation of the option with respect to other goals. See in this handbook, [Chap. 45, EU Risk Regulation and the Uncertainty Challenge](#).

## Risk as an Attribute

A risk is a chance that something bad will happen. Most people are averse to risk. They will forgo a benefit to decrease risk. For example, they will pay premiums to insure a house against damage from fire. If a fire occurs, having insurance reduces the net harm the fire causes the homeowner and so reduces risk by reducing the stakes. Another way to reduce risk is to lower the probability of harm. A homeowner may do this by installing a fire alarm and keeping a fire extinguisher handy. These measures lower the probability of serious fire damage and, by lowering that probability, reduce the risk of that harm.

Economists take the shape of a person’s utility curve for a commodity to represent a person’s aversion to risk. In a typical graph of a utility curve for money, the horizontal axis represents money, and the vertical axis represents utility. If a person is averse to risks concerning money, then the person’s utility curve for money is concave down, that is, it bends down, as [Fig. 20.1](#) shows.



**Fig. 20.1**  
**A concave utility curve for money**

If a person's utility curve for money is concave, then the more money she has, the less she wants an additional dollar. As a result, she will not buy a gamble at its fair monetary value. Take a gamble that yields a dollar gain or a dollar loss depending on whether a fair coin lands with heads up or with heads down. The gamble's fair monetary value is \$0. Given the concavity of the utility curve for money, a dollar lost has more value than a dollar gained. Therefore, the expected utility of the gamble is less than the utility of \$0. So the gamble's utility for the person is less than the utility of no monetary gain or loss. Hence, she refuses the gamble when it is offered for free.

Keeney and Raiffa ([1976]1993, pp. 148–158) and Gollier (2001) explain the view that aversion to risk amounts to having a concave utility curve for commodities. See also in this handbook, [Chap. 5, The Economics of Risk: A \(Partial\) Survey](#).

The multi-attribute approach to risk, as Keeney and Raiffa ([1976]1993) present it, does not treat risk as an attribute. It assumes that the shape of the utility curve for an option's possible outcomes registers aversion to risk, so that representing risk as an attribute and assigning utilities to risks is unnecessary. However, this section argues that the shape of the utility curve for an option's possible outcomes does not adequately represent aversion to risk. The section therefore departs from that representation of aversion to risk and advocates treating risk as an attribute in a multi-attribute assessment of a risky option.

Aversion to risk taken as the concavity of a utility curve for a commodity does not distinguish the aversion from the diminishing marginal utility of the commodity. It also makes aversion to risk relative to a commodity and so not an attitude that applies to risks involving multiple commodities. Finally, the account of aversion to risk neither describes risk nor explains aversion to risk. It is silent about the nature of risk and the origin of aversion to risk.

Treating risk as an attribute of an option conforms with the understanding of risk in ordinary life. Risk is a feature of some options, and aversion to that feature of options, risk aversion, prompts aversion to risky options. More specifically, an option is risky because of

a chance that it will produce a bad result. Aversion to risk is aversion to that chance, the risk of a bad result. It is more intense the greater the risk, and a risk is greater the greater the probability of a bad result and the worse the bad result. In a wager, risk depends on the stakes as well as the probability of losing the stakes. Bell and Raiffa (1988) call this ordinary type of aversion to risk “intrinsic risk aversion” and distinguish it from diminishing marginal utility.

This chapter, for theoretical simplicity, extends risk to include the chance that something good will happen. An agent may be averse to having a good result depend on chance. Aversion to risk in the broad sense includes such an aversion. It amounts to aversion to having things one cares about, both good and bad, depend on chance.

An option’s risk depends not only on the probabilities of the option’s possible outcomes but also on their utilities. A probability distribution of the utilities of possible outcomes displays the option’s risk, but some measures of risk try to represent the risk more succinctly. They claim that some parameters of the distribution capture the essence of the option’s risk. A rough measure of an option’s risk takes it to depend on the variance of the utilities of the option’s possible outcomes. Often, the larger the variance, the greater the risk. A gamble that yields a gain of \$10 or a loss of \$10 depending on the outcome of a fair coin toss is riskier than a gamble on the toss that yields a gain of \$1 or a loss of \$1. The variance of the utility of the possible outcomes is greater for the first gamble than for the second gamble, even though the probabilities of winning and of losing are the same for both gambles. The variance is a crude measure of risk, however, and theorists such as Keeney and Raiffa ([1976]1993, pp. 159–165) and Luce (1980) advance more subtle measures. See in this handbook, [Chap. 17, The Mismeasure of Risk](#).

Allais’s paradox presents a good reason for treating risk as an attribute rather than letting the shape of utility curves handle it. In one version of this paradox, a person prefers \$3,000 to a 4/5 chance of \$4,000 and prefers a 1/5 chance of \$4,000 to a 1/4 chance of \$3,000. These preferences are rational, but no utility function for money explains their rationality.  $U(\$3,000) > 4/5U(\$4,000)$  according to the first preference, and  $1/5U(\$4,000) > 1/4U(\$3,000)$  according to the second preference. No utility function for money yields these inequalities, however, because multiplying both sides of the second inequality by 4 yields the inequality  $4/5U(\$4,000) > U(\$3,000)$ , which is the opposite of the first inequality. The two inequalities are inconsistent. The shape of the utility curve for money therefore does not explain the two preferences. Rather, aversion to risk accounts for the first preference, and a judgment that two low probabilities are not significantly different accounts for the second preference. This explanation of the two preferences treats risk as an attribute of risky options.

Risk depends on probabilities, and because probabilities may be either physical or evidential, risks may be either physical or evidential. A physical probability depends on physical features of the world. The physical probability of drawing a red ball from an urn depends on the percentage of red balls in the urn. An evidential probability for a person depends on the person’s evidence. For a juror, the evidential probability that a defendant is guilty fluctuates as the prosecuting and defending attorneys present evidence for and against the defendant’s guilt. The defendant’s physical probability of guilt does not fluctuate during the trial; it is either 0% or 100% throughout the trial. At the start of a typical trial, the juror knows the evidential probability but not the physical probability of guilt. Evidential probabilities for a person are accessible to the person because they depend only on the evidence in the person’s possession. Evidential risks are also called epistemic risks. For more discussion of them, see this handbook, [Chap. 18, Unreliable Probabilities, Paradoxes, and Epistemic Risks](#).

The size of an option's evidential risk depends not only on the probabilities and utilities of the option's possible outcomes but also on the weight of the information yielding the possible outcomes' probabilities, that is, the extent of the evidence grounding the probability assignment to possible outcomes. The information's weight affects the probability assignment's robustness. The greater the weight, the more robust the probability assignment. A robust probability assignment, an assignment unlikely to change much as new information arrives, reduces risk. If two drugs have the same risk of a side effect but one has been more extensively tested, then that drug, because of the more extensive evidence grounding probabilities concerning its possible outcomes, carries less evidential risk. A risk-averse person prefers taking the more extensively tested drug because it carries less evidential risk.

## Risk-Return Assessment of Investments

---

A common application of the multi-attribute approach to risk is the risk-return method of assessing investments. The method evaluates an investment by evaluating the risk it involves, by evaluating the investment's expected return, and by combining the two evaluations to obtain an overall evaluation of the investment. The method treats risk as an attribute of an investment and treats aversion to risk as an aversion to that attribute; it does not take an investor's aversion to risk as the concavity of the investor's utility curve for money.

A financial planner may use a questionnaire to assess an investor's attitude toward risk and toward return, and may use financial data to obtain an investment's risk and expected return. Investing in bonds has less risk but a lower expected return than investing in stocks. Because of an investor's aversion to risk and willingness to reduce returns to reduce risk, a financial planner may recommend investing in bonds rather than in stocks. The reduction in risk compensates for the lower expected return.

To assess an investment's expected return and risk, a financial planner typically uses data about the investment type's past returns and their variation. Financial experts debate measures of risk and aversion to risk, and the method of combining aversion to an investment's risk with attraction to its expected return to obtain an overall evaluation of the investment. A common *mean-risk* formula begins with the investment type's mean return and subtracts from it a proportion of the standard deviation (the square root of the variance) for returns from the investment type. That is, from expected return it subtracts risk. Raiffa (1968, pp. 55–56) presents and criticizes this mean-risk formula. Despite flaws, the formula may be useful for rough appraisals of investments in a limited range of cases. It fits the mold for a multi-attribute analysis of an investment's utility if the mean return's utility replaces the mean return, the risk's utility replaces the risk's size, and the constant of proportionality weights the risk's utility so that the mean return's utility minus the proportion of the risk's utility equals the investment's utility.

Another method of evaluating investments uses the *coefficient of variation*:  $\sigma/\mu$ , where  $\sigma$  is the standard deviation for returns from the investment type, and  $\mu$  is the mean of returns from the investment type, or the investment's expected return. The smaller the coefficient of variation, the better the investment. This risk-return method of evaluation uses division rather than subtraction to combine assessments of an investment with respect to the two attributes, return and risk. It fits the multi-attribute mold if the utility of the mean return divided by the utility of the risk's size is inversely proportional to the investment's utility.

A refinement of the coefficient of variation is the *Sharpe ratio*:  $(R - R_f)/\sigma$ , where  $R$  is the investment type's expected return,  $R_f$  is the risk-free rate of return, and  $\sigma$  is the standard deviation for returns from the investment type. The factor  $R - R_f$  is expected return in excess of return obtainable without risk, or *excess expected return*. The Sharpe ratio, as the previous two formulas, uses standard deviation to measure risk. It uses division to combine assessments of an investment with respect to excess expected return and with respect to risk. The larger the ratio, the better the investment. The ratio yields a multi-attribute assessment of an investment if the utility of the expected excess return divided by the utility of the risk's size is proportional to the investment's utility.

The risk-return methods of assessing investments are useful even if they do not yield investments' utilities but only numbers that agree with a utility ranking of investments. This section uses risk-return evaluations of investments only to illustrate the multi-attribute approach to risk and so does not explore issues concerning limitations and refinements of this method of evaluating investments. Weirich (1987) reviews some of the issues. Textbooks on financial management, such as Brigham and Houston (2009, Chap. 8), also explore them.

## Multi-Attribute Utility Functions

---

A multi-attribute utility function obtains an option's utility from the utilities of the option's attributes. Its justification shows that the option's utility depends exclusively on the utilities of its attributes so that options that are equivalent in utilities of attributes have the same utility. Addition is commonly, but not exclusively, used for combining attributes' utilities to obtain an option's utility. A justification for using addition includes showing that attributes' utilities have equal weights from the perspective of the option's utility. If they do not have equal weights, then an option's utility may equal, instead of the sum of its attributes' utilities, a weighted sum of their utilities.

Whether an option's utility is a function of its attributes' utilities depends partly on the set of attributes. An option's utility may be a function of its attributes' utilities for some set of attributes but not for another set. This section assumes a set of attributes that meet the criteria that section  [Government Regulation](#) reviewed. Meeting those criteria enhances the prospects that an option's utility is a function of its attributes' utilities.

An aggregation function yielding an option's utility from its utilities with respect to attributes implies an ability to move from one option to another of the same utility by replacing an instance of an attribute of the first option with another instance of the same attribute, provided that the two instances of the attribute have the same utility. In fact, an option's utility is a function of its attributes' utilities if and only if interchange of equivalent parts of options yields equivalent options. The possibility of such interchanges assumes that multiple instances of an attribute with the same utility exist and that substituting in an option one instance of an attribute for another instance yields distinct options. Consequently, interchanges do not occur for the attribute of financial gain if the utility of money is strictly monotonically increasing. In that case, no two financial gains have the same utility. No interchanges of distinct gains with the same utilities are possible. Moreover, if all attributes have strictly monotonically increasing utility, then an option's utility is a function of its attributes' utilities. No two options with attributes having the same utilities have different utilities. If options differ, then they

have different instances of attributes, and those instances have different utilities. This section treats cases in which two options may have equivalent attributes so that the existence of an aggregation function is not trivial.

Additive aggregation functions require that attributes' utilities have some properties of independence. For example, increasing an attribute's utility increases an option's utility and increases it by the same amount whatever are the other attributes' utilities. A nonadditive aggregation function may impose some related properties of independence. According to the function, attributes' utilities may have independent effects on options' utilities. A change in utility with respect to an attribute may produce the same effect on an option's utility no matter how other attributes are fixed. If the method of aggregation is multiplicative, then one attribute's utility is proportional to the option's utility given that other attributes' utilities are held fixed. According to a type of independence that section [Aggregation Functions](#) examines, as an attribute's utility increases, the option's utility increases no matter how other attributes are fixed.

This section focuses on additive aggregation functions because of their prominence in the literature. It considers conditions sufficient for using a sum or weighted sum to represent preferences among options, but argues that an additive utility representation of preferences is not enough to establish genuine additivity of utilities. The section examines first a set of conditions for an additive representation that a theorem of Harsanyi (1955) suggests. The conditions involve coherence of preferences and responsiveness of some preferences to other preferences.

Harsanyi's theorem is about personal and social preferences. Preferences, either personal or social, are *coherent* if and only if they may be represented as maximizing expected utility. Social preferences are *Pareto* if and only if they are responsive to the unanimous preferences of society's members, more precisely, if and only if they conform to the traditional Pareto principle that requires a society to prefer one option to another if all members prefer the first, or some members are indifferent between the options and other members prefer the first. Broome (1991, p. 160) states Harsanyi's theorem this way: "Suppose that each person has coherent preferences, and that social preferences are coherent and Paretoian. Then social preferences can be represented by an expectational utility function that is the sum of expectational utility functions representing the individuals' preferences." An *expectational utility function* is one that takes an option's utility to equal the expected utility of the option's possible outcomes.

Although Harsanyi's theorem treats personal and social preferences, because its proof is formal, it is open to alternative interpretations that replace personal and social preferences with other types of preferences. This section applies Harsanyi's theorem to aggregation of a person's restricted utility functions for options, taken with respect to attributes, to obtain the person's comprehensive utility function for options. Because aggregating personal preferences to form social preferences structurally resembles aggregating a person's restricted preferences to form her comprehensive preferences, the theorem governs aggregation to obtain an option's utility as well as aggregation to obtain an option's social utility.

The definition of coherent preferences applies straightforwardly to an agent's preferences among options. Preferences among options with respect to an attribute are coherent if and only if preferences among options that are equal in all other attributes, and offer chances for various instances of that attribute, may be represented as maximizing expected utility. An agent's

preferences among options are “Paretian” if and only if the agent prefers one option to another given that she prefers the first option to the second with respect to every attribute, or prefers the first option to the second with respect to some attributes and is indifferent between the options with respect to the other attributes.

The theorem implies that if preferences are coherent for each attribute and for options, and if preferences for options follow unanimous preferences with respect to attributes, or more precisely are “Paretian,” then an option’s utility may be represented as a weighted sum of its attributes’ utilities.

Broome uses the traditional probability agreement theorem to object to the usual application of Harsanyi’s theorem. He states the probability agreement theorem this way (p. 160): “Suppose that each person has coherent preferences. Then social preferences cannot be both coherent and Paretian, unless everyone agrees about the probability of every state of nature.” Broome holds that the probability agreement theorem refutes the assumptions of Harsanyi’s theorem because people do not agree about probabilities. Putting the point less contentiously, the probability agreement theorem limits applications of Harsanyi’s theorem to ideal cases involving agents who agree about probabilities. In any case, the probability agreement theorem is no obstacle to this section’s application of the theorem. A single probability assignment governs all of a person’s restricted preferences. So Harsanyi’s theorem applies to an agent’s restricted and comprehensive preferences concerning options.

Does Harsanyi’s theorem, as this section applies it, therefore offer a rationale for additive aggregation of an option’s restricted utilities, taken with respect to attributes, to obtain the option’s comprehensive utility? Keeney and Raiffa ([1976]1993, pp. 527–528) and Ellis (2006, n. 19, n. 22) point out this rationale for additivity. A close look at the theorem and additivity, however, shows that the theorem does not entail the additivity of a rational ideal agent’s utility assignments.

Harsanyi’s theorem is a representation theorem. It shows that preferences meeting certain conditions have a certain type of representation; utility functions with a certain structure may be defined to represent the preferences. The theorem does not show that restricted utility functions stating degrees of desire, when added using weights, yield a comprehensive utility function stating degrees of desire, but only that utility functions representing restricted and comprehensive preferences can be constructed so that adding the functions representing restricted preferences, using weights for them, yields the function representing the comprehensive preferences. Thus, taking utility as section [Multi-Attribute Utility Analysis](#) does, Harsanyi’s theorem does not justify using addition, under an assignment of weights to attributes, to aggregate an option’s utilities with respect to attributes into the option’s comprehensive utility.

The theorem, nonetheless, has an important consequence about the measurement of utility. Assuming that additive aggregation is correct, the theorem justifies inferring attribute weights from preferences among options and preferences among options with respect to attributes. Representation of these preferences generates a function that obtains an option’s utility in the representation as a weighted sum of its utilities with respect to attributes in the representation. The additive function that the representation generates is unique given choices of utility scales for options and attributes. Therefore the function uses the same weights as the additive function that goes from an option’s genuine utilities with respect to attributes to the option’s genuine utility, given that those utilities use the same scales that utilities in the representation use.

Also, although Harsanyi's theorem does not justify addition of utilities with respect to attributes, it justifies comparability of attributes in the cases it treats. Because of Harsanyi's theorem, coherent preferences among options with respect to each attribute, and coherent and "Pareto" preferences among options, entail that the preferences have an additive representation that assigns weights to the attributes and so compares the attributes. Such a representation does not exist if the attributes are incomparable. Given incomparable attributes, a rational ideal agent does not form a complete set of preferences among options and among options with respect to each attribute. As the theorem shows, forming the preferences despite the incomparability results in defective preferences. Section 2 Incomparable Attributes treats rational choice given incomparable attributes and incomparable options.

Some famous representation theorems establish conditions necessary and sufficient for an additive representation of preferences involving multiple attributes. These representation theorems assume in the background a rich set of preferences involving a set of attribute types.

Suppose that  $X_i$  is an attribute,  $x_i$  is a value of the attribute,  $u$  is a utility function for an option taken as a combination of attribute values, and  $u_i$  is a utility function over values of  $X_i$ . Debreu (1960) showed that given attributes  $X_1, X_2, \dots, X_m$ ,  $n \geq 3$ , an additive utility function  $u(x_1, x_2, \dots, x_n) = \sum_i u_i(x_i)$  exists if and only if the attributes are mutually preferentially independent. The attributes are *mutually preferentially independent* if and only if every subset is preferentially independent of its complementary set. A set of attributes  $Y$  is *preferentially independent of its complementary set*  $-Y$  if and only if for some  $z'$ , if  $(y', z') \geq (y', z')$ , then  $(y', z) \geq (y'', z)$  for all  $z, y', y''$ , where  $\geq$  is the relation of *weak preference*, that is, preference or indifference.

Gorman (1968) showed how to reduce the work needed to establish mutual preferential independence. Let  $Y$  and  $Z$  be subsets of the attribute set  $S = \{X_1, X_2, \dots, X_n\}$  such that  $Y$  and  $Z$  overlap, but neither is contained in the other, and such that the union  $Y \cup Z$  is not identical to  $S$ . If  $Y$  and  $Z$  are each preferentially independent of their respective complements, then  $Y \cup Z$ ,  $Y \cap Z$ ,  $Y - Z$ ,  $Z - Y$  and  $(Y - Z) \cup (Z - Y)$  are each preferentially independent of their respective complements. Because preferential independence among some sets entails preferential independence among other sets, showing mutual preferential independence does not require examining every subset and its complement.

Fishburn (1965) established a necessary and sufficient condition for an additive representation when the objects of preferences are lotteries over bundles of attributes. The attributes  $X_1, X_2, \dots, X_n$  are *additive independent* if and only if preferences over lotteries on  $X_1, X_2, \dots, X_n$  depend only on their marginal probability distributions and not on their joint probability distribution. Additive independence is necessary and sufficient for an additive representation. This result is related to Harsanyi's theorem about conditions sufficient for an additive representation, as Keeney and Raiffa ([1976]1993, pp. 527–528) explain.

The representation theorems begin with a set of preferences and present conditions under which the preferences have a representation using an additive utility function. A corresponding normative principle requires utilities to be additive. The conditions for an additive representation do not establish a normative principle of additivity, just as Harsanyi's conditions for an additive representation do not establish a normative principle of additivity. A normative principle of additivity for utilities concerns utilities defined independently of additivity, for example, defined as rational degrees of desire as in section 2 Multi-Attribute Utility Analysis. That preferences have an additive representation does not show that the preferences arise from utilities complying with the normative principle.

Take a decision problem in which an agent has just two salient options, chocolate ( $c$ ) and vanilla ( $v$ ), and he prefers the first. Using the set of attributes, richness ( $r$ ) and smoothness ( $s$ ), the preference has an additive representation. For example, defining utilities as follows works:  $U(c) = U_r(c) + U_s(c) = 3 + 2$  and  $U(v) = U_r(v) + U_s(v) = 1 + 3$ . This representation yields the preference  $c > v$ . However, suppose that the utility function representing the preference does not accurately report the agent's degrees of desire. According to an accurate utility function,  $U_r(c) + U_s(c) = 3 + 2$  and  $U(v) = U_r(v) + U_s(v) = 2 + 3$ . The preference for chocolate holds because the caffeine in the chocolate breaks the tie. Then although the preference has an additive representation using the two attributes, richness and smoothness, the preference does not arise because of addition of utilities with respect to those two attributes. The utilities that the representation introduces are not the utilities that the agent assigns, and the set of attributes omits a relevant consideration.

This example does not fall among the cases that the representation theorems treat; it does not meet the theorems' background assumptions. The theorems apply to a rich set of preferences and not to a single preference. Nonetheless, such examples show that the representation theorems do not establish a normative principle stating that the utility of an option is a sum of its utilities with respect to the attributes in a set, even if they present necessary and sufficient conditions for preferences having an additive utility representation. The step from additive representations to genuine additivity needs argumentation.

## Intrinsic-Utility Analysis

---

Intrinsic-utility analysis divides an option's utility into the utilities of its attributes. The attributes are realizations of basic intrinsic desires and aversions. This section explains these basic intrinsic attitudes and shows how they ground an analysis of an option's utility. It draws on Weirich (2001, Chap. 2). The main principle of intrinsic-utility analysis asserts that when a rational agent knows an option's outcome, the option's utility is a sum of the intrinsic utilities of the objects of the basic intrinsic attitudes realized given the option's realization. The principle is normative because it governs rational degrees of desire. It derives from the additivity of the intrinsic utilities of realizations of basic intrinsic attitudes.

The difference between intrinsic and extrinsic desire arises because of desire's variable scope. Some desires hold all things considered, and others hold given a restricted range of considerations. A person may desire that a proposition obtain because of the proposition's total outcome, including its causal consequences, or because of the proposition's logical consequences alone. In the first case the desire is extrinsic, and in the second case it is intrinsic. A person typically has an intrinsic desire for pleasure and an extrinsic desire for money. Aversion similarly may be either intrinsic or extrinsic.

An intrinsic desire may rest on other intrinsic desires. For example, an intrinsic desire for two pleasures may rest on intrinsic desire for the first pleasure and an intrinsic desire for the second pleasure. An intrinsic desire that does not rest on any other intrinsic desire is basic. A basic intrinsic aversion has an analogous definition. A person's aversion to a risk is apt to be a basic intrinsic aversion.

A proposition's utility for a person is a rational degree of desire that the proposition obtain. Varying the scope of considerations on which desire depends produces different kinds of utility. Intrinsic utility measures desire arising from considerations of limited scope. A proposition's

intrinsic utility for a person is the person's degree of intrinsic desire that the proposition hold. This equals the person's degree of desire that the proposition hold considering only the proposition's logical consequences. Intrinsic utility contrasts with ordinary, comprehensive utility, which unlike intrinsic utility measures desire arising from considerations of unlimited scope. A proposition's ordinary, comprehensive utility for a person is the person's degree of desire that the proposition hold all things considered, including the proposition's causal consequences as well as its logical consequences.

A person's degrees of desire, if rational, meet standards that take account of the person's abilities. This chapter treats ideal agents whose rational degrees of desires meet the highest standards and consequently satisfy principles of utility, such as consistency: if two propositions are logically equivalent, then their utilities are equal. The principles of utility vary with the type of utility. Comprehensive, but not intrinsic, utility obeys the expected utility principle requiring an option's utility to equal the option's expected utility. Intrinsic utility, because it rests on logical consequences, does not respond to chances, and so the expected utility principle does not govern it.

In this section, a basic intrinsic attitude is either a basic intrinsic desire or a basic intrinsic aversion. A proposition that specifies for each basic intrinsic attitude whether it is realized characterizes the relevant aspects of a possible world. It represents the possible world. The proposition's intrinsic utility equals its ordinary, comprehensive utility because the proposition specifies everything that matters. Its logical consequences include all its relevant consequences.

Suppose that an agent is rational and ideal. For the agent, a world's intrinsic utility is a sum of the intrinsic utilities of the objects of the basic intrinsic attitudes that the world realizes. The reasons for this principle of addition are roughly the following. The realization of a basic intrinsic attitude makes the same contribution to the intrinsic utility of any world that realizes the basic intrinsic attitude. Conjunction of objects of basic intrinsic attitudes is a concatenation operation for intrinsic utility, and addition of intrinsic utilities represents that operation. An intrinsic utility assignment therefore is additive with respect to basic intrinsic attitudes. A proposition's intrinsic utility is the sum of the intrinsic utilities of the objects of basic intrinsic attitudes whose realizations the proposition entails.

For example, suppose that an agent has basic intrinsic desires for health and for wisdom and has no other basic intrinsic attitudes. Suppose also that the intrinsic utility of health is 3 and the intrinsic utility of wisdom is 2. Then the intrinsic utility of health and wisdom is 5.

Next, suppose that an agent has a basic intrinsic desire for each of two pleasures during successive intervals. She also has a basic intrinsic desire for continuity of pleasure during the two intervals. Then the intrinsic utility of both pleasures is the sum of the intrinsic utilities of the two pleasures and also the intrinsic utility of pleasure's continuity during the two intervals. The conjunction of the propositions that each pleasure obtains during its interval entails realization of those propositions and also realization of the proposition that pleasure is continuous during the two intervals.

Intrinsic utility grounds a multi-attribute approach to risk because an aversion to risk typically is intrinsic and basic. Intrinsic-utility analysis therefore authorizes dividing the intrinsic utility of a risky option's possible outcome, a world that might be realized if the option were realized, into the intrinsic utility of the option's risk and the intrinsic utilities of other objects of basic intrinsic attitudes that would be realized if the possible outcome were realized. An option's ordinary, comprehensive utility equals the utility of the proposition that the option's world obtains. Given certainty, an agent knows the option's world, the possible

outcome that obtains, and an option's utility equals its world's intrinsic utility. Given uncertainty, an option's utility is the expected value of the intrinsic utility of the proposition that the option's world obtains, that is, a probability-weighted average of the intrinsic utilities of the option's possible outcomes.

## Trustee Decisions

---

Multi-attribute approaches to risk, with risk as an attribute, facilitate decisions in which a trustee decides for a client, for example, decisions in which a physician decides for a patient about a treatment for an illness. This section examines trustee decisions in which the trustee is an expert and a client authorizes the trustee to use expert information to advance the client's basic goals.

The trustee decides for the client, this section assumes, by maximizing utility among options, using the trustee's information and the client's goals. This procedure is appropriate when the trustee's information includes all the client's relevant information and the objective is a decision that the client might make, if rational, given possession of the trustee's expert information.

To implement the procedure, the following steps are effective. For an option, the trustee supplies a probability assignment to possible outcomes, and the client supplies a utility assignment to possible outcomes. The client's utility assignment to a possible outcome depends on his assessment of the option's attributes, including the risk the option generates. The client's assessment of the option's risk is uninformed, and the trustee replaces it with an informed assessment of the risk. Then the trustee recalculates the possible outcome's utility, substituting her informed assessment of the risk for the client's uninformed assessment, but using the client's assessments of the option's other attributes. Finally, the trustee assigns utilities to options applying the expected utility principle with her probability assignment and the recalculated utilities of possible outcomes, and identifies an option of maximum utility in the set of options.

Suppose that a physician has two treatment options that offer a patient equal chances of recovery. Accordingly, the patient takes the options to generate equal risks. The utility of a treatment's possible outcome depends on the attributes of risk and of health. Adding more attributes is easy, but using just these two attributes keeps the illustration simple. Because the first treatment promises slightly faster recovery, its possible outcomes have slightly greater utilities than their counterparts with the second treatment. Using for possible outcomes the physician's probability assignment and the client's utility assignment, the first treatment's utility is superior. However, the physician knows that the second treatment has undergone more extensive testing than has the first treatment. Hence, the second treatment generates less risk than does the first treatment. Substituting the physician's assessments of the treatments' risks, utilities for the second treatment's possible outcomes are slightly greater than for their counterparts with the first treatment. The second treatment's utility is therefore greater than the first treatment's utility. Thus, using expert assessments of risk in the multi-attribute calculation of possible outcomes reverses utility maximization's recommendation.

To illustrate, suppose that  [Table 20.2](#) represents the patient's utilities for the two treatments given success and also given failure. Each cell represents a treatment's possible outcome. The first number in a cell evaluates the possible outcome with respect to risk, and the second evaluates it with respect to health, including recovery time.

Assuming that a possible outcome's utility equals the sum of the utilities of its attributes, **Table 20.3** shows the utilities of each treatment's possible outcomes.

Because the first treatment's utilities are higher in each column, their probability-weighted sum is greater than the probability-weighted sum of the utilities of the second treatment's possible outcomes. Therefore the first treatment maximizes expected utility, and so utility.

The comparison reverses, however, after substituting informed assessments of risks in place of the patient's assessments. **Table 20.4** shows the change.

**Table 20.5** displays utilities of possible outcomes after recalculating their utilities from their informed utilities with respect to the attributes of risk and of health.

Using informed assessments of risks and so informed utilities of possible outcomes, the second treatment has greater expected utility, and hence greater utility, than has the first treatment. The physician should therefore recommend the second treatment to the patient. In this example, informed assessments of risks reverse utility maximization's original recommendation.

**Table 20.2**  
**Patient's assessments of attributes**

	Success	Failure
First treatment	2, 11	2, 6
Second treatment	2, 10	2, 5

**Table 20.3**  
**Utilities of outcomes**

	Success	Failure
First treatment	13	8
Second treatment	12	7

**Table 20.4**  
**Informed assessments of attributes**

	Success	Failure
First treatment	2, 11	2, 6
Second treatment	4, 10	4, 5

**Table 20.5**  
**Informed utilities of outcomes**

	Success	Failure
First treatment	13	8
Second treatment	14	9

This example assumes, as section [Risk as an Attribute](#) argues, that risk is an attribute of risky options and their possible outcomes. Granting this, multi-attribute utility analysis is plainly a valuable tool for trustee decisions. In particular, it assists government regulatory decisions of the sort section [Government Regulation](#) reviewed. In those decisions the government agency is a collective trustee, and the public is the client that the agency serves. The agency uses its expert information to assess risks, and it uses its informed assessments of risks to make decisions on the public's behalf.

## Further Research

---

Multi-attribute approaches to risk are still under construction. Research on the approaches tackles unfinished tasks. Among the tasks to be completed are detailed justifications of the approaches and generalizations of the approaches so that they cover realistic and not just idealized decision problems.

## Aggregation Functions

---

The justification of a multi-attribute approach to risk derives an option's utility from the utilities of the option's relevant attributes. The utilities of the option's attributes are comparable if an option's utility is a function of its attributes' utilities, and each attribute's utility affects the option's utility. A typical multi-attribute assessment of an option's utility entails the comparability of the utilities of the option's relevant attributes. So its justification requires establishing their comparability. Section [Incomparable Attributes](#) treats comparability, and this section treats other issues pertaining to justification of multi-attribute assessments of options.

In some decision problems, it is convenient to compare options using selected attributes if the options are alike with respect to other attributes. For example, a financial planner may compare options alike in risk by comparing their expected returns. A multi-attribute utility function justifies this simplification if utilities of attributes are separable. Separability is a matter of independence of attributes' utilities. The utilities of attributes are *separable* if and only if holding constant all attributes but one while increasing the utility of that attribute increases an option's utility. Varian (1984, Sect. 3.14) calls this feature of attributes' utilities *functional separability*.

Separability rules out a certain type of complementarity between attributes, that is, cases in which increasing utility with respect to one attribute while holding other attributes constant lowers an option's utility. Such complementarity holds between wealth and security, for instance, if, assuming poor security measures, increasing wealth lowers comprehensive utility because it makes kidnapping more likely.

Traditional utilitarianism treats the well-being of each individual in a group as an attribute affecting collective utility. It takes collective utility to be a function of the utilities of the individuals' levels of well-being, or, more briefly, utilities for individuals. Utilities for individuals are separable, in fact, additively separable, according to utilitarianism. That is, adding utilities for individuals yields collective utility.

The utility of equality raises an objection to utilitarianism. According to the objection, an option that promotes equality in a group benefits the group but does not distribute a benefit to each member of the group. Consequently, increasing a member's utility while holding other members' utilities constant may lower collective utility by destroying equality in the group. This lowering of collective utility is contrary to the separability that utilitarianism advances. A defense of utilitarianism argues that a group's members benefit from equality so that destroying equality by raising utility for one member also lowers utility for the other members. Because utilities for the other members change during the lowering of collective utility, that lowering is compatible with separability. Hence, the utility of equality does not discredit a utilitarian separation of individuals' utilities.

Value holism argues that the moral value of a whole does not equal the sum of the moral values of its parts. Suppose that an individual suffers. Adding to the situation a second individual who takes pleasure in the first's suffering makes the situation worse, not better, although pleasure is a good. This case counts against the separability of moral values, because increasing pleasure while holding suffering constant decreases total value. The counterexample works, however, only if all pleasures count as good. Value theory may distinguish types of pleasure and deny that all are good. Because pleasure at another's suffering is bad, its lowering the situation's value does not refute the separability of moral values.

These examples from moral theory illustrate attacks on and defenses of separability. Multi-attribute utility theory's reliance, in some applications, on the separability of attributes' utilities faces similar attacks and may marshal similar defenses.

Multi-attribute utility theory does not claim that every set of attributes offers a way of deriving an option's utility. It selects attributes that permit the derivation. Also, it selects the function that combines attributes' utilities to obtain an option's utility. It need not select an additive function, or even a function that makes utilities of attributes separable. A multi-attribute utility analysis needs only one set of attributes and one function that together permit deriving an option's utility from its attributes' utilities. Because of an analysis's modest requirements, the prospects of justifying a multi-attribute utility analysis of an option's utility are good in many cases.

## Realism

---

This chapter adopts several idealizations to facilitate presentation of multi-attribute approaches to risk. For example, it assumes that agents are ideal and so have no cognitive limits. It also assumes that decision problems are ideal so that every option has a utility with respect to each attribute and also has a comprehensive utility; attributes and options are compared, and comparisons are precise.

Realistic agents in realistic decision problems cope with cognitive limits, imprecision, and incomparability. Theorists seek to remove idealizations and extend to realistic cases multi-attribute approaches to risk. Perhaps some idealizations are indispensable and so impose limits on managing risk by weighing pros and cons. Nevertheless, the multi-attribute approach to risk should extend as far as those limits allow. It should relax as many idealizations as possible. For discussion of realistic decision principles and their limits, see in this handbook, [Chap. 21, Real-Life Decisions and Decision Theory](#).

Consider the idealization that an agent in a decision problem assigns, for every option, a probability and a utility to each possible outcome with respect to every attribute. This idealization is strong even for a rational agent with ample resources for reflection. Such an agent may lack evidence that warrants a precise probability assignment or may lack experience that warrants a precise utility assignment. For a tourist in Paris, a rational assignment of probability to rain tomorrow may be imprecise, and so may a rational assignment of utility to visiting the Louvre tomorrow. Even given full rationality and ample computational resources, the tourist may fail to remove all imprecision in probability and utility assignments.

In some decision problems imprecision does not prevent identifying an optimal option. Perhaps no matter the precise probability of rain tomorrow and no matter the precise utility of visiting the Louvre tomorrow, that visit is optimal for a tourist in Paris. Other options have less utility under every refinement of probabilities and utilities. A sensitivity analysis may establish the robustness of the visit's position at the top of a ranking of options.

A topic for further research is whether an option's optimality under some refinement of probabilities and utilities suffices for the option's rationality. Good (1952, p. 114) advances the view that it suffices, but Elga (2010) challenges that position, arguing that given unsharp probabilities it condones incoherent sequences of decisions, for example, rejecting each of a pair of gambles that together guarantee a gain.

Some idealizations concern the options in a decision problem. In a standard problem some option has maximum utility, and the agent has a stable basis for comparing options. The agent has stable basic goals, and assuming an option's realization does not alter the basis of its comparison with other options. Moreover, weights for attributes and utilities of possible outcomes are independent of both the option realized and the state of the world that obtains.

To handle the dependence of weights and utilities on options and states, utility theory may take an option's possible outcomes to be possible worlds. Then possible outcomes include everything an agent cares about. Their utilities, because inclusive, are independent of options and states. The main drawback of the broad view is that possible outcomes, if comprehensive, do not occur in multiple contexts. That complicates measurement of their utilities.

Game theory generates decision problems in which assuming an option's realization alters the option's comparison with other options. Consider a two-person game with a unique Nash equilibrium, that is, an assignment of strategies to agents such that each agent's strategy is a best response to the other agent's strategy. An agent's supposition that she departs from her Nash strategy may carry information that her opponent departs from his Nash strategy. That information affects the utility of her departure from her Nash strategy.

Because decisions in games have special features, the principle of straightforward utility maximization may need refinement to handle those decisions. For example, it may give way to the principle of ratification. That principle recommends an option that maximizes utility on the assumption that the option is realized, as Weirich (2007, Chap. 8) explains.

The next section considers relaxing the idealization that attributes and options are compared. It examines a generalization of the principle of utility maximization designed to accommodate decision problems with uncomparable options.

## Incomparable Attributes

---

For a person, two objects of desire or aversion are incomparable if the person cannot compare the two. For example, a connoisseur may like both music and wine yet be unable to compare the two. The connoisseur neither prefers one to the other, nor is indifferent between the two, and, moreover, despite ample time for reflection, cannot form a preference or become indifferent between them. In a decision problem, if a person has options with incomparable attributes, then the options may be incomparable, too. If the options are incomparable, then the ranking of options is incomplete. Choosing from the top of the ranking may be unwarranted.

Helzner (2009) argues that multi-attribute assessments of options should accommodate incomplete orderings of options. In some decision problems, for two options, the agent neither prefers one option to the other, nor is indifferent between them. The agent's preference is indeterminate. Indeterminacy of preferences between options may arise from incomparable attributes of options. It may also arise because of indeterminacy in the weighting of attributes, that is, the incommensurability of the attributes. Two attributes may be comparable but not commensurable because although the first is weightier than the second, it is not weightier by a precise amount.

An agent in a decision problem may be at fault for not having a determinate preference between a pair of options. Perhaps he should have reflected and formed a preference. In some cases, however, a failure to form a preference is excused. This section takes it for granted that a rational ideal agent may face a decision problem in which the relevant attributes of options are incomparable or incommensurable so that for good reason the agent does not compare all options. It considers a method of evaluating the agent's decision despite an incomplete ranking of options.

Not all decision problems in which options have incomparable attributes create complications. In some decision problems, options are comparable although attributes are not. Suppose that a decision problem has two salient options, and the first is better than the second with respect to every attribute. Then the first is better than the second and makes a rational choice. The options' comparison does not require that the attributes be comparable. Incomparable attributes trouble evaluations of decisions only when the attributes' incomparability blocks comparison of options.

Also, in some decision problems a preference ranking of options need not be complete to be useful. Suppose that an option is at the top of the agent's ranking of options although the ranking omits some options available in the decision problem. A plausible decision principle permits that option because the agent prefers no other option. The permission is uncontroversial if the ranking omits only one option because it is incomparable with some options in the ranking, and the agent prefers the option at the top of the ranking to the option that the ranking omits.

This section targets decision problems in which incomparable attributes create incomparable options that make the preference ranking of options incomplete. The ranking omits an option because it is incomparable to some options in the ranking. However, no rival in the incomplete ranking defeats it. In particular, the agent does not prefer the option at the top of the ranking.

For example, a new engineer may have two job offers, and may use salary and location as attributes to evaluate the offers. Suppose that for the engineer the two attributes are incomparable. She cannot compare for either job its salary with its location and so cannot settle the attributes' relative weights and the job's overall utility. Suppose also that the job with the greater

salary has a less attractive location than has the other job, and the engineer cannot settle whether the greater salary compensates for the less attractive location. Incomparability of attributes blocks multi-attribute assessments of options. The engineer does not assign a utility to either job, and the jobs are incomparable. Her ranking of options omits them, but no option in the ranking defeats them. Resolving such decision problems requires generalized decision principles. Can multi-attribute utility theory formulate principles that handle such decision problems?

Etzioni (1986) introduces multiple utility scales to handle incomparable attributes that do not fit on the same scale. He proposes two utility scales, one for pleasure and one for morality, but is open to adding other utility scales. He rejects combining an option's assessments with respect to the two utility scales to form the option's comprehensive utility, but does not explain how to resolve decision problems without combining the two sorts of assessment of an option.

A general problem with using multiple utility scales to handle incomparable attributes is that eventually an agent must move from an assessment of options using attributes to a choice. Suppose that in a decision problem, instead of putting morality and pleasure on the same scale, a decision procedure puts morality on one scale and pleasure on another scale. It does not compare morality and pleasure. Then it aggregates an option's utilities on the two scales to obtain an overall evaluation of the option. Its aggregation of utilities with respect to attributes is unjustified because it does not compare attributes. Incomparability thwarts justified aggregation of utilities with respect to attributes and thus blocks the standard multi-attribute approach to decisions.

How do agents resolve decision problems despite such obstacles? One possibility is that an agent attends to only one utility scale in a decision problem, and that scale directs the agent's decision. Sometimes an agent attends to morality, and other times to pleasure. This possibility moves away from evaluative to descriptive decision theory. It treats an agent's motivation rather than an agent's justification for a decision. This chapter, which treats evaluative decision theory, puts aside that method of resolving a decision problem with incomparable options.

A second possibility is that an agent uses a lexicographical ordering of options to reach a decision. Perhaps morality has absolute priority over pleasure. Pleasure's role may be limited to breaking ties between options that are morally permissible.

The lexicographic method of using attributes to reach a decision takes an option's choiceworthiness to be a nonadditive function of its evaluation with respect to each attribute. It dispenses with options' utilities, the mainstay of the standard multi-attribute approach to decisions. This revised version of the multi-attribute approach assigns utilities to options with respect to attributes but does not assign comprehensive utilities to options. It uses utilities with respect to attributes to reach decisions without the intermediary of options' comprehensive utilities and without comprehensive comparisons of options. As in the example, it ranks options according to a primary attribute and then breaks ties using a secondary attribute.

Levi (1986, pp. 28–34, 80–82) describes two-step and multi-step decision procedures that use options' comparisons with respect to attributes. He considers a two-step procedure that begins with an agent's identification of options that are best according to at least one of the agent's principal values. In the second step, from these options, the agent selects an option that is best according to a tie-breaking, secondary value. For example, suppose that an employer ranks one job candidate best according to education and another candidate best according to experience. If education and experience are the primary values, then those two candidates are the finalists. The employer may then choose between the two finalists according to their performance during an interview, a secondary value.

A problem confronts this two-step procedure. In the example, suppose that a third candidate is second according to education and also experience, best according to interview, and so best all things considered. The decision procedure fails to select that candidate despite her overall optimality. If options have an overall ranking, a decision procedure ought to use it. For the two-step decision procedure to be plausible, it must be restricted to cases in which options lack an overall ranking.

On the other hand, given the absence of an overall ranking, the decision procedure seems too strict. In some cases, it requires an agent to select an option even though the agent does not prefer it to a rival option. For example, it requires the employer to select the candidate who gave the better interview even though the employer, without any failure of reflection, does not prefer that candidate to the others. It seems that if rationality does not require an agent to prefer one option to all others, then it does not require the agent to choose that option.

A common alternative to lexicographic methods is less demanding and also more broadly applicable. It applies to cases in which options are incomparable, not only all things considered, but also considering just a single attribute. For example, it applies to the employer's decision problem even if the job candidates are incomparable with respect to education, experience, and performance during an interview.

Suppose that in a decision problem two options in contention are incomparable. Rather than insist on an option whose utility is at least as great as any other option's utility, either overall or with respect to some attribute, the alternative principle declares that each incomparable option is a rational choice. Does rationality in fact permit either option? Is the permissive decision principle correct?

To begin answering, consider a crucial difference between incomparability and indifference. If an agent is indifferent between two options, adding a small benefit to one option breaks the tie. This is not so with incomparable options. If  $x$  and  $y$  are incomparable options, so are  $x+$  and  $y$ , where  $x+$  is a small improvement of  $x$ . This disanalogy between indifference and incomparability makes the permissive decision principle condone incoherent sequences of choices, Peterson (2007) argues.

Suppose that a safety measure has incomparable attributes such as a reduction in wages and a reduction in risk to health. The same holds for inaction. As a result, a decision about the safety measure surveys incomparable options. The same holds for subsequent decisions about similar safety measures. Following the permissive decision principle may produce an incoherent sequence of decisions about safety measures.

To show this, consider options  $x$ ,  $x+$ , and  $y$ . Suppose that rationality, assuming the permissive decision principle, permits the following three choices. (1) The agent may choose option  $x$  over option  $y$ , because  $x$  is better than  $y$  with respect to one attribute and worse than  $y$  with respect to another, incomparable attribute so that the options are incomparable. (2) The agent may choose option  $x+$  over option  $x$ , because  $x+$  is slightly better than  $x$  with respect to one attribute and is the same as  $x$  with respect to other attributes. (3) The agent may choose option  $y$  over option  $x+$ , because, as with  $x$  and  $y$ , the options are incomparable. Then rationality condones a cycle of choices from  $y$  to  $x$ , from  $x$  to  $x+$ , and from  $x+$  to  $y$ .

The cycle makes the agent a money pump, Peterson (2007) observes. A person ready to make the choices in the cycle may be led through the cycle many times, each time paying a sum to make the switch from  $x$  to  $x+$  until he is bankrupt. The possibility of a money pump makes

a case against the rationality of the choices in the cycle and thus against the permissive decision principle. The principle condoning any choice among incomparable contenders is mistaken, Peterson concludes.

Some ways of defending the permissive decision principle are available, however. A method of preventing incoherent choices given incomparable options is for an agent to keep a record of past choices and evaluate current choices in light of past choices so that an aversion to incoherent choices blocks cycles. For instance, an agent after choosing  $x$  over  $y$  and  $x+$  over  $x$  should revise evaluation of  $y$  and  $x+$  so that they are no longer incomparable. In light of past choices, the agent should prefer  $x+$  to  $y$  because of an aversion to cyclical choices, and to the particular cycle realized by a choice of  $y$  over  $x+$ .

A standard rebuttal is that sunk costs, or in this case past choices, do not count. All that counts is the current utility assignment to options. This rebuttal is not decisive. A rejoinder contends that past choices count if an agent is averse to cyclical choices. Past choices settle whether a current choice completes a cycle of choices. Given an aversion to cycles, past choices influence current utility assignments to options. The choices lower an option's current utility assignment if because of them the option's realization completes a cycle. Because of an aversion to cycles, it is rational to have a utility assignment to options that is attuned to a current option's consequences given past choices. Moreover, it is rational to have an aversion to cyclical choices. In fact, rationality requires such an aversion in cases where cyclical choices lead to sure losses. So a rational ideal agent following the permissive decision principle does not make an incoherent sequence of choices.

Because the permissive decision principle does not lead to incoherent sequences of choices, it is defensible. It gives multi-attribute approaches to risk a way of handling incomparable attributes.

## Outlook

Despite unresolved issues, multi-attribute approaches to risk have attractive features, such as the transparency they bring to assignments of utilities to options. Because of these attractive features, many decision theorists are motivated to refine multi-attribute approaches to meet challenges rather than to abandon these approaches in the face of challenges. Decision theorists' continuing investigations give multi-attribute approaches to risk a bright future.

## References

- Bell D, Raiffa H (1988) Marginal value and intrinsic risk aversion. In: Bell D, Raiffa H, Tversky A (eds) *Decision making*. Cambridge University Press, Cambridge, pp 384–397
- Brigham E, Houston J (2009) *Fundamentals of financial management*, 12th edn. South-Western Cengage Learning, Mason
- Broome J (1991) *Weighing goods*. Blackwell, Oxford
- Debreu G (1960) Topological methods in cardinal utility theory. In: Arrow K, Karlin S, Suppes P (eds) Mathematical methods in the social sciences. Stanford University Press, Stanford, pp 16–26
- Elga A (2010) Subjective probabilities should be sharp. *Philosophers' Imprint* 10(5):1–11, [www.philosophersimprint.org/010005/](http://www.philosophersimprint.org/010005/)
- Ellis S (2006) Multiple objectives: a neglected problem in the theory of human action. *Synthese* 153:313–338
- Etzioni A (1986) The case for a multiple-utility conception. *Econ Philos* 2:159–183

- Fishburn P (1965) Independence in utility theory with whole product sets. *Oper Res* 13:28–45
- Gollier C (2001) The economics of risk and time. MIT Press, Cambridge, MA
- Good IJ (1952) Rational decisions. *J R Stat Soc B* 14:107–114
- Gorman WM (1968) The structure of utility functions. *Rev Econ Stud* 35:367–390
- Harsanyi J (1955) Cardinal welfare, individualist ethics, and interpersonal comparisons of utility. *J Polit Econ* 63:309–321
- Helzner J (2009) On the application of multi-attribute utility theory to models of choice. *Theory Decis* 66:301–315
- Keeney R, Raiffa H ([1976]1993) Decisions with multiple objectives: preferences and value tradeoffs. Cambridge University Press, Cambridge, pp 527–528
- Kleindorfer P, Kunreuther H, Schoemaker P (1993) Decision sciences. Cambridge University Press, Cambridge
- Levi I (1986) Hard choices. Cambridge University Press, Cambridge
- Luce RD (1980) Several possible measures of risk. *Theory Decis* 12:217–228
- Mullen J, Roth B (1991) Decision-making: its logic and practice. Rowman and Littlefield, Savage
- Peterson M (2007) On multi-attribute risk analysis. In: Lewens T (ed) Risk: a philosophical view. Routledge, London, pp 68–83
- Raiffa H (1968) Decision analysis: introductory lectures on choices under uncertainty. Addison-Wesley, Reading
- Varian H (1984) Microeconomic analysis, 2nd edn. Norton, New York
- Weirich P (1987) Mean-risk decision analysis. *Theory Decis* 23:89–111
- Weirich P (2001) Decision space: multidimensional utility analysis. Cambridge University Press, Cambridge
- Weirich P (2007) Realistic decision theory: rules for nonideal agents in nonideal circumstances. Oxford University Press, New York



# 21 Real-Life Decisions and Decision Theory

*John R. Welch*

Saint Louis University, Madrid Campus, Madrid, Spain

<b><i>Introduction</i></b> .....	<b>546</b>
<b><i>History</i></b> .....	<b>547</b>
The Maximizing Objection .....	547
The Precision Objection .....	550
<b><i>Current Research</i></b> .....	<b>552</b>
Preliminaries .....	552
The Binary Case .....	558
The Finite General Case .....	562
<b><i>Further Research</i></b> .....	<b>563</b>
Transitivity .....	563
Independence .....	565
Decision Rules .....	571

**Abstract:** Some decisions result in cognitive consequences such as information gained and information lost. The focus of this study, however, is decisions with consequences that are partly or completely noncognitive. These decisions are typically referred to as “real-life decisions.” According to a common complaint, the challenges of real-life decision making cannot be met by decision theory. This complaint has at least two principal motives. One is the maximizing objection that to require agents to determine the optimal act under real-world constraints is unrealistic. The other is the precision objection that the numeric requirements for applying decision theory are overly demanding for real-life decisions. Responses to both objections are aired in the section [History](#) of this chapter. The maximizing objection is addressed with reference to work by Weirich and Pollock, while the precision objection is countered via a proposal by Kyburg and another by Gärdenfors and Sahlin. However, the section [Current Research](#) urges a different response to the precision objection by introducing a comparative version of decision theory. Drawing on Chu and Halpern’s notion of generalized expected utility, this version of decision theory permits many choices to be based on merely comparative plausibilities and utilities. Finally, the section [Further Research](#) undertakes an open-ended exploration of three of the assumptions upon which this form of decision theory (and many others) is based: transitivity, independence, and plausibilistic decision rules.

## Introduction

---

Decisions can be classified by their consequences. The consequences of some decisions are cognitive, such as information gained or information lost; the consequences of other decisions are noncognitive, such as economic gain, political embarrassment, esthetic enjoyment, legal imbroglio, or military advantage. Where the consequences of concern to the agent are purely cognitive, the agent confronts a *cognitive decision*; where the situation is mixed, with relevant consequences that are both cognitive and noncognitive, the agent must make a *partly cognitive decision*; and where the relevant consequences are purely noncognitive, the agent faces a *noncognitive decision*.

The concept of a *real-life decision* is both irremediably vague and undeniably useful. Rather than attempt to define the concept in abstract terms, I will simply note that the contrast class for real-life decisions appears to be cognitive decisions (despite the fact that cognitive decisions are, strictly speaking, just as real and just as much a part of life as partly cognitive and noncognitive decisions). Hence most real-life decisions are either partly cognitive or noncognitive. Examples of real-life decisions can be suggested by questions such as these: Whom should I vote for? Would it be better to cancel this trip? Must I tell the truth to my nosy neighbor? Do I need a second medical opinion? How should I invest this windfall?

Decision theory has been dogged by complaints that it is inapplicable to real-life decisions. There appear to be two standard objections. The maximizing objection takes issue with the decision-theoretic directive to maximize, that is, to choose the optimal act. Though this anti-maximizing approach can take different forms (Elster 1979; Levi 1986; Slote 1989), the best-known appears to be Herbert Simon’s advocacy of satisficing (Simon 1982). Given the complexity of our environment and our limitations in gathering and processing information, the goal of maximizing is unattainable, according to Simon. Instead, we should aim to satisfice by seeking results that are merely good enough relative to some threshold of expected utility. The second complaint about decision-theoretic realism is the precision objection: decision

theory exacts numerical precision that we can very rarely supply. Strict Bayesian decision theorists take the probability and utility functions that underlie expected utilities to determine sharp numeric values (de Finetti 1937; Savage 1972). But real-life decision makers usually operate without nearly as much numeric data; they are characterized, in fact, by numeric poverty. Hence, the objection goes, decision theory cannot be applied to most real-world decisions.

This chapter explores the prospects for obtaining decision-theoretic guidance for real-life decisions. It proceeds in three stages. The section [History](#) reviews several perspectives on the maximizing and precision objections. The section [Current Research](#) presents the author's proposal for meeting the precision objection. Finally, the section [Further Research](#) offers suggestions on how this proposal might be solidified by future investigation.

In keeping with this volume's focus on risk, I will concentrate on decisions under risk, that is, decisions for which the decision maker can estimate the probability (or plausibility) of states of the world relevant to the decision. I will say nothing further about decisions under ignorance, that is, decisions for which the decision maker is unable to make such probabilistic (or plausibilistic) determinations.

## History

### The Maximizing Objection

We begin with two well-known perspectives on maximizing. Despite areas of agreement, Paul Weirich and John L. Pollock hold positions that are fundamentally opposed. Weirich claims that rationality requires maximization; Pollock demurs.

Weirich's defense of maximization unfolds against the backdrop of a theory of idealizations. An idealization simplifies by focusing on some explanatory factors and excluding others. In physics, for example, a theory of motion may idealize by excluding the explanatory factor of air resistance – in effect, assuming it is zero.

Historically, decision theory has developed by employing a full complement of idealizations. Weirich's *Realistic Decision Theory* identifies and classifies these idealizations (2004, Chap. 3). It then performs a sort of controlled demolition by subtracting them one by one. Idealizations of the agent are eliminated in Chaps. 4–7; idealizations of the situation, in Chaps. 8–9. With each jettisoned idealization, decision theory becomes both more realistic and more general.

What turns out to be a lengthy and complex process can be illustrated by a few examples from its inception. Weirich initially assumes agents who are perfect and fully informed. Such godlike agents would not engage in deliberation and decision, for “they can maximize utility spontaneously” (2004, p. 22). Specifically, they would maximize informed utility by complying with the *principle of utility maximization for acts*: “Among the acts you can perform at a time, perform one whose utility is at least as great as any other’s utility” (2004, p. 20).

Removing the idealization of perfect agents but retaining that of full information leaves us with imperfect agents for whom it would make sense to deliberate and compare utilities. For these agents, the foregoing principle for acts is generalized as the *principle of utility maximization for decisions*: “Among the decisions you can make at a time, make one whose utility is at least as great as any other’s utility” (2004, p. 23).

Casting off the idealization of full information carries us one step further in the direction of realism and generality. Agents who are uncertain of the outcomes of acts under consideration must consider the probabilities of states of the world as well as the utilities of outcomes. They do this by taking an act's utility to be its expected utility, that is, a weighted average of the act's outcome utilities in which the relevant state probabilities are the weights. Suppose that an agent equipped with a probability function  $\mu$  and a utility function  $v$  is considering the performance of an act  $a$ . Relevant to the decision are a finite number of attributable states of the world  $s_1, s_2, \dots, s_n$  with corresponding state probabilities  $\mu(s_1, e), \mu(s_2, e), \dots, \mu(s_n, e)$  based on the evidence  $e$ . In addition, performance of  $a$  when exactly one of  $s_1, s_2, \dots, s_n$  obtains would produce outcomes  $o_1, o_2, \dots, o_m$ , respectively, and these outcomes have utilities  $v(o_1), v(o_2), \dots, v(o_n)$ . The expected utility  $E$  of  $a$  given  $e$  can then be defined as follows:

$$E_{a,e} = \sum_{i=1}^n v(o_i)\mu(s_i, e).$$

For imperfect, partially informed agents, the principle of utility maximization for decisions would be generalized as the *principle of expected utility for decisions*: “Among the decisions you can make at a time, make one whose expected utility is at least as great as any other’s expected utility.”

The three maximization principles cited in the three preceding paragraphs are all treated in Chap. 2 of *Realistic Decision Theory*. But the process of stripping away idealizations continues throughout the work until Weirich finally proposes a *principle of comprehensive rationality*: “A rational decision maximizes self-conditional utility among self-supporting options with respect to a quantization of conscious beliefs and desires after a reasonable effort to form and become aware of relevant beliefs and desires, acquire pertinent *a priori* knowledge, and correct unacceptable mistakes” (2004, p. 191). Technical terms such as “quantization,” “unacceptable mistakes,” “self-supporting options,” and “self-conditional utility” will not be defined here. Readers interested in further detail are advised to consult the relevant sections of the work (2004, pp. 69–70, 119–123, 147–154, and 155–158, respectively). Even without these definitions, however, it is evident that Weirich remains committed to a highly general form of maximization.

Pollock’s *Thinking about Acting* (2006) rejects the goal of maximization. Its point of departure is the thesis that theories of ideal decision makers tell us little about how real decision makers, whether humans or machines, should make decisions. Theories of ideal decision makers are theories of *warranted choice*, that is, choices that “would be justified if the agent could complete all possibly relevant reasoning.” By contrast, theories of real decision makers are theories of *justified choice*, choices that “a real agent could make given all the reasoning it has performed *up to the present time* and without violating the constraints of rationality” (Pollock 2006, p. 6). Because of cognitive limitations, real decision makers often make choices that are justified but not warranted.

In Pollock’s view, subjective expected utility theory is a theory of ideal rationality. One of its principal departures from realism is the optimality (or maximizing) principle: choose optimal (or maximal) solutions to problems. Pollock objects to it on a number of grounds. In classical decision theory, the optimality principle is formulated in terms of actions: choose the action that maximizes expected utility. But Pollock argues that actions must be chosen as parts of plans; hence the basic alternatives faced by decision makers are plans, not actions. If we attempt

to reformulate the optimality principle in terms of plans, Pollock raises three further objections: “Although plans are evaluated in terms of their expected utilities, they cannot be chosen on the basis of a pairwise comparison because plans can differ in scope.... [Furthermore,] there are always potentially infinitely many plans that compete with a given plan. An agent cannot be expected to survey them all, so it is unrealistic to expect the agent to choose optimal plans. There may not even be optimal plans. For every plan, there may be a better plan” (2006, p. 167). In short, alternative plans may not be comparable because they differ in scope; the injunction to choose an optimal plan from an infinity of possible plans is unrealistic; and optimal plans may not even exist.

What Pollock proposes instead is a theory of real rationality that he calls “locally global planning.” Normal planning is local planning – a plan for getting to Vienna, say. But local plans can be merged to create a global or master plan – to write a definitive book on the *Wiener Kreis*, for example. Local plans are often just good enough to get the agent started, but they can be improved in the attempt to implement them. Even in their improved versions, however, they can be expected to be less than optimal. Yet this does not mean that they are less than rational, for the rationality of real decision makers consists in continually bettering these plans. In short, “a rational decision maker should be an *evolutionary planner*, not an *optimizing planner*” (2006, p. 187).

Pollock recognizes the kinship between locally global planning and satisficing. But he distinguishes them as follows: “Satisficing consists of setting a threshold and accepting plans whose expected utilities come up to the threshold. The present proposal requires instead that any plan with a positive expected utility is defeasibly acceptable, but only defeasibly. If a better plan is discovered, it should supplant the original one. Satisficing would have us remain content with the original” (2006, p. 187 note 4). In Pollock’s view, locally global planning is more adaptable, more responsive to changing circumstances, than satisficing.

Though the debate between maximizers and anti-maximizers rages on (Byron 2004), this is no place for an exhaustive treatment. Nevertheless, since the  Current Research section’s proposal for applying decision theory to real-life decisions is a maximizing proposal, I feel obliged to comment very briefly on Pollock’s critique of maximizing.

The point that some plans cannot be pairwise compared because they have different scopes is well-taken. Someone who is thinking about attending either a 2-year community college or a 4-year college is not considering plans of equal scope. I would suggest, however, that the plans can be compared in either of two ways. One is to supplement the community college plan with whatever additional plan the agent might consider as an alternative to the third and fourth years in a 4-year college. When packaged with this additional plan, the community college plan can be compared to the 4-year college plan with an eye to maximizing. The other approach is to compare only the first and second years at a 4-year college with 2 years at a community college, looking once again for the optimal plan. Evidently, these maneuvers can be adapted to other plans with incongruent scopes.

Pollock also charges that the injunction to choose the optimal plan is unrealistic because “there are always potentially infinitely many plans that compete with a given plan.” The key point here is that a plan is optimal relative to other plans. For an ideal agent, these other plans constitute an infinite set. For a real agent, however, the other plans form a finite subset of the infinite set (cf. Giere 1985, p. 87; Weirich 2004, p. 142). Granted, then, that since a real agent lacks an ideal agent’s ability to evaluate an infinite number of plans, it would be unrealistic to expect her to choose the optimal plan relative to this set. But it would not be unrealistic to

expect her to choose the optimal plan relative to the finite subset she can consider. To fend off this charge of unrealism, then, we only need to ensure that “optimal plan” is realistically understood.

Still, Pollock contends, there may not be optimal plans. A well-known example is Leonard J. Savage’s problem of choosing your income to be some figure less than \$100,000 per year (Savage 1972, p. 18). Since there are an infinite number of options, each with an alternative that has greater utility, there is no optimal plan. As Savage points out, however, the supposition of an infinite number of options requires “abstracting from the indivisibility of pennies”; that is, any system of currency will divide the income continuum between \$0 and \$100,000 into a finite number of intervals, exactly one of which is optimal. But if we ignore the wet blanket of realism and simply postulate an infinite number of options, none of which is optimal, what would that mean for a strategy of maximization? Weirich responds that “The [maximization] rule needs generalization, not correction” (2004, p. 144). In such cases, maximization should be generalized to satisficing. The principle of utility maximization for decisions cited above would become the *principle of utility satisfaction for decisions*: “Among the decisions you can make at a time, make one whose expected utility is satisfactory.” In other words, satisfice.

Note, however, that Weirich’s use of satisficing is different than Simon’s (Weirich 2004, p. 145). Simon proposes satisficing as a decision procedure, whereas Weirich uses it (like the more specific concept of utility maximization) as a standard of evaluation. A decision procedure is, as it were, first person: to make this decision, I will apply this rule. But a standard of evaluation may be first, second, or third person: anyone at all can evaluate a decision by appealing to this standard.

Though no one will confuse the foregoing remarks with a vindication of maximization, I hope that they suffice to show that the maximizing strategy remains a live option. That, at least, is my intention, and I propose to employ this strategy in the quest for decision-theoretic realism below. In order to do this, however, we first need to investigate alternative grounds for the claim that decision theory is unrealistic.

## The Precision Objection

---

As noted in the section [Introduction](#), classical decision theory assumes probability and utility functions that determine sharp numeric values. This is an idealization, however (Weirich 2004, Chaps. 4 and 5); real-life decisions must usually be made with far less numeric information. Inevitably, then, decision theorists face the objection concerning precision: decision theory demands numbers that real-life decision makers rarely possess.

The present subsection canvasses two responses to this objection: one by Henry E. Kyburg, the other by Peter Gärdenfors and Nils-Eric Sahlin. We can set the stage for both responses with a remark by Savage: “One tempting representation of the unsure is to replace the person’s single probability measure  $P$  by a set of such measures, especially a convex set” (1972, p. 58 note +). A convex set of probability measures ensures that, for any probability measures  $p$  and  $p'$  that belong to the set and for  $0 \leq \alpha \leq 1$ ,  $\alpha p + (1 - \alpha)p'$  also belongs to the set. For example, for a state  $s$  where  $p(s) = 0.2$ ,  $p'(s) = 0.4$ , and  $\alpha = 1/2$ ,  $(\alpha p + (1 - \alpha)p')(s) = 0.3$ . In other words, a convex set of probability measures defines a probabilistic interval. Though convex sets of probability measures may not always be appropriate (Kyburg 2006), one approach to the problem of numeric poverty is to retrench by dropping back from point-valued to interval-valued functions.

Drawing on earlier work on interval-valued probability functions (Kyburg 1961; Good 1962; Levi 1974), Kyburg proposes that expected utility is interval valued as well (1979). The gist of the proposal can be conveyed by a simple numeric example. Suppose that the only states relevant to a decision to perform act  $a_1$  or act  $a_2$  are  $s_1$  and  $s_2$ . If the probability of  $s_1$  is the closed interval (0.25, 0.50), then the probability of  $s_2$  is the closed interval (0.75, 0.50). Assume that the outcome utility of performing  $a_1$  when  $s_1$  obtains is 0.80 and the outcome utility of performing the same act when  $s_2$  obtains is 0.40. Then the expected utility of  $a_1$  is:

$$(0.25 \times 0.80, 0.50 \times 0.80) + (0.75 \times 0.40, 0.50 \times 0.40) = (0.50, 0.60).$$

The expected utility of  $a_2$  would be an interval as well.

The expectation intervals for  $a_1$  and  $a_2$  will either overlap or not. If they do not, the principle of maximizing expected utility would require that the act associated with the rightmost interval be chosen. If the intervals do overlap, on the other hand, the coincidence is either total or partial. If total, there is no decision-theoretic reason to choose one act over the other; they are equally choiceworthy. But if the coincidence is partial, the maximizing principle breaks down. In these cases, Kyburg suggests the adoption of an additional rule such as minimax or maximax to complete the decision (Kyburg 1979, p. 434).

Consistent with Savage's remark above, Gärdenfors and Sahlin regard the strict Bayesian assumption that decision makers' beliefs about states can be represented by a single probability measure as unrealistic. They choose to rely on a set  $P$  of epistemically possible probability measures instead (Gärdenfors and Sahlin 1982). They do not insist that  $P$  be convex, though they do discuss the convex special case where probability measures establish probabilistic intervals (1982, pp. 365–366). In this respect, then, their approach is more general than Kyburg's.

Even though Gärdenfors and Sahlin utilize a set of probability measures to represent decision makers' beliefs about states, they claim that not all such beliefs can be represented in this way. In particular, beliefs about the epistemic reliability of the probability measures in  $P$  elude the representation. To illustrate this point, take a spectator who is considering bets on the outcomes of three tennis matches. In the first, the spectator is extremely well informed about the two players and regards them as evenly matched; in the second, the spectator has only heard that one of the players is far superior to the other without hearing which player this is; and in the third, the spectator knows nothing whatever about the tennis skills of the two players. Gärdenfors and Sahlin point out that, relative to the spectator's information, each player has a 50% chance of winning. But the epistemic reliability of these probabilities is different in each case. Specifically, her probability measure for the first match is much more reliable than the probability measures for the other two matches. Gärdenfors and Sahlin propose to take epistemic reliability into account by specifying a real-valued measure  $\rho$  of the epistemic reliability of probability measures.

To do this, they describe a two-step decision procedure. The first step is to restrict the set of epistemically possible probability measures to a subset of epistemically reliable probability measures. Let  $\rho_0$  represent the decision maker's desired level of epistemic reliability. Then  $P/\rho_0$ , the set of epistemically possible probability measures  $P$  given the level  $\rho_0$  of epistemic reliability, is the subset of epistemically reliable probability measures. The members of this set are the probability distributions that the decision maker actually takes into account. The second step of the procedure is to apply the maximin criterion for expected utilities: Choose the act with the highest minimal expected utility. That is, the decision maker calculates the

expected utility of each act for each probability measure in  $P/\rho_0$ , identifies the minimal expected utility for each act, and chooses the act with the highest minimal expected utility.

## Current Research

---

Like the previous section, the present section seeks a more realistic decision theory. In fact, it takes its cue from the interval-based approaches just reviewed. Interval decision theory retrenches by dropping back from point-valued to interval-valued functions. The approach to be floated here attempts to retrench even further by dropping back from interval to comparative functions. The issue to be addressed is the possibility of comparative decision theory: Could decision theory be applied with merely comparative values for probabilities and utilities?

I will suggest an answer to this question in three stages. First of all, the principal assumptions to be employed are laid out in the subsection [● Preliminaries](#). These assumptions are then invoked to develop a comparative approach to binary choices between acts. Finally, this comparative approach is extended to choices among any finite number of acts. Throughout, the focus will be on the partly cognitive and noncognitive decisions characteristic of real-life decision making.

### Preliminaries

---

Due to space limitations, I will introduce the main assumptions underlying the theory baldly, with little or no justification. Only the final assumptions concerning decision rules and relative disutility will require fuller initial treatment. Three of these main assumptions, however, will receive further scrutiny in the section on [● Further Research](#).

*Propositions:* Decision-theoretic acts, states, and outcomes can be thought of as propositions (Jeffrey 1983, pp. 82–85).

*Finitude:* Because real-life decision makers can consider only a finite number of items at a given decision point, acts, states, and outcomes are assumed to be finite (cf. Gärdenfors and Sahlin 1982, p. 364; Weirich 2004, pp. 24, 28, 142).

*Transitivity:* If  $a$  is strictly preferred to  $b$  and  $b$  is strictly preferred to  $c$ , then  $a$  is strictly preferred to  $c$ . (Transitivity is treated in more detail in the first part of the section on [● Further Research](#).)

*Independence:* Let act  $a$  have outcomes  $x$  and  $y$  and act  $b$  have outcomes  $x$  and  $z$ . Then the choice between  $a$  and  $b$  should be made by ignoring the common outcome  $x$  (cf. Maher 1993, pp. 10, 12, 63–83). That is, the decision should be made independently of  $x$ . (Independence is discussed more fully in the second part of the section on [● Further Research](#).)

*Plausibility:* A probability measure may map propositions about states to numbers in the unit interval. But real-life decision makers are often unable to provide reliable numeric probabilities for states. Consequently, I will work with the more general notion of plausibility. A plausibility measure can map propositions about states to members of any partially ordered set (Friedman and Halpern 1995). Though plausibility values can be restricted to the special case of the unit interval (e.g., Klir 2006, p. 166), they can also include qualitative values like high, likely, and impossible provided they are partially ordered. In one influential conception

(Friedman and Halpern 1995), a plausibility function  $\pi$  returns values that are bounded by nonnumeric limits  $\top$  and  $\perp$ , where  $\top$  represents the maximum plausibility and  $\perp$  the minimum plausibility. So conceived,  $\pi$  satisfies the following requirements for propositions  $q$  and  $r$  (cf. Chu and Halpern 2004, pp. 209–210):

- Pl1. If  $q$  is contradictory,  $\pi(q) = \perp$ .
- Pl2. If  $q$  is tautologous,  $\pi(q) = \top$ .
- Pl3. If  $q$  implies  $r$ ,  $\pi(q) \leq \pi(r)$ .

Comparing Pl1–Pl3 to the standard Kolmogorov axioms for probability shows that probability is a special case of plausibility.

*Order:* Classic formulations of decision theory assume that beliefs, desires, and preferences for acts have representations that are totally ordered: either  $a$  is strictly preferred to  $b$ , or  $b$  is strictly preferred to  $a$ , or  $a$  and  $b$  are equally preferred. Even von Neumann and Morgenstern were uneasy about this assumption: “It is very dubious, whether the idealization of reality which treats this postulate as a valid one, is appropriate or even convenient” ([1944]1953, p. 630; cf. Aumann 1962, p. 446; Ok 2002; Ok et al. 2004). I take it to be a simple fact that agents are sometimes unable to reasonably determine whether one utility, plausibility, or preference for an act is greater than, equal to, or less than another. In the interest of realism, therefore, I assume that utilities, plausibilities, and preferences for acts can be incomparable.

To express relations of order in a realistic way, the nonstrict comparative term “ $\preceq$ ” will be drafted as primitive. The  $\preceq$  relation establishes a partial order; that is, it is reflexive, antisymmetric, and transitive.

In the context of plausibility, “ $\preceq$ ” can be read as “is less plausible than or equally plausible to” or “is not more plausible than.” The following plausibility relations are straightforwardly definable in terms of it, conjunction (“ $\wedge$ ”), and negation (“ $\neg$ ”):

$$\begin{aligned} \text{Infraplausibility } [\pi(s_1, e) < \pi(s_2, e)] &=_{\text{df}} [\pi(s_1, e) \preceq \pi(s_2, e)] \wedge -[\pi(s_2, e) \preceq \pi(s_1, e)] \\ \text{Supraplausibility } [\pi(s_1, e) > \pi(s_2, e)] &=_{\text{df}} -[\pi(s_1, e) \preceq \pi(s_2, e)] \wedge [\pi(s_2, e) \preceq \pi(s_1, e)] \\ \text{Equiplausibility } [\pi(s_1, e) = \pi(s_2, e)] &=_{\text{df}} [\pi(s_1, e) \preceq \pi(s_2, e)] \wedge [\pi(s_2, e) \preceq \pi(s_1, e)] \\ \text{Incomparability } [\pi(s_1, e) \mid \pi(s_2, e)] &=_{\text{df}} -[\pi(s_1, e) \preceq \pi(s_2, e)] \wedge -[\pi(s_2, e) \preceq \pi(s_1, e)]. \end{aligned}$$

The primitive “ $\preceq$ ” will be used in different settings, and its context will determine its sense. In addition to the plausibilistic usage just described, “ $v(o_1) \preceq v(o_2)$ ” can be read as “the utility of outcome  $o_1$  is no greater than the utility of outcome  $o_2$ ” and “ $PE(a_1) \preceq PE(a_2)$ ” as “the plausibilistic expectation of act  $a_1$  is no greater than the plausibilistic expectation of act  $a_2$ ” (plausibilistic expectation is defined just below). With these nonstrict relations as primitives, relations of utility and plausibilistic expectation analogous to infraplausibility, supraplausibility, equiplausibility, and incomparability can be easily defined.

*Decision rules:* To have a single, all-purpose decision rule would be ideal. Unfortunately, the prospects for finding one are not very good. The principal reason for this is the hard fact of incomparability. Situations in which plausibilities or utilities are incomparable are so different from situations in which both plausibilities and utilities are comparable that I think we are unlikely to find a single rule that suits all occasions. However, the prospects are good, I think, for finding a suite of decision rules that cover the spectrum of comparable and incomparable data. In fact, it is possible to find three cognate decision rules to carry out this task. These three rules will be introduced in the following paragraphs.

Before these rules are stated, I will sketch the general framework that relates them to each other. Savage famously showed that preferences for one act over another can be represented by

comparative relations between expected utilities (1972, pp. 69–82). Savage's representation theorem has a plausibilistic generalization, as Francis Chu and Joseph Halpern have demonstrated (2008, pp. 12–13). To sketch this result, we assume states of the world  $s_1, s_2, \dots, s_n$  and corresponding state plausibilities  $\pi(s_1, e), \pi(s_2, e), \dots, \pi(s_n, e)$  based on the evidence  $e$ . We also assume that performing act  $a$  when exactly one of  $s_1, s_2, \dots, s_n$  obtains would produce outcomes  $o_1, o_2, \dots, o_n$ , respectively, and that these outcomes have utilities  $v(o_1), v(o_2), \dots, v(o_n)$ . In addition, we rely on Chu and Halpern's notion of generalized expected utility. This generalization of expected utility is defined for an expectation domain  $D = (U, P, V, \oplus, \otimes)$ , where  $U$  is a set of utility values ordered by a reflexive binary relation  $\leq_u$ ;  $P$  is a set of plausibility values ordered by a binary relation  $\preceq_p$  that is reflexive, antisymmetric, and transitive;  $V$  is a set of expectation values ordered by a reflexive binary relation  $\leq_v$ ; the multiplication-like operation  $\otimes$  maps  $U \times P$  to  $V$ ; and the addition-like operation  $\oplus$  maps  $V \times V$  to  $V$  (2004, pp. 209–211; 2008, pp. 6–10). Then the generalized expected utility  $GEU$  of  $a$  given  $e$  can be expressed as:

$$GEU_{a,e} = \bigoplus_{i=1}^n v(o_i) \otimes \pi(s_i, e).$$

Chu and Halpern show that, where  $\leq_A$  is a preference relation over a set  $A$  containing a finite number of acts  $a_1, a_2, \dots, a_n$ ,

$$a_1 \leq_A a_2 \text{ iff } GEU(a_1) \leq GEU(a_2)$$

That is, preferences for acts can be represented by comparative relations between generalized expected utilities.

The decision rule associated with  $GEU$  is to maximize generalized expected utility. As Chu and Halpern point out, this rule is universal in the following sense: it establishes the same ordinal rankings as any decision rule that satisfies a trivial condition. The condition is that the rule weakly respect utility, that is, that act preferences track outcome utilities for all constant acts (2004, pp. 216, 219, 226–227). Constant acts are constant in the sense that their outcomes do not depend on states of the world (Savage 1972, p. 25).

For the purposes of real-life decision making, I propose to adopt three special forms of Chu and Halpern's decision rule. The principal rule is the first; the other two cover very special cases. To introduce this first rule, let  $D = (U, P, T, V, \oplus, \otimes)$  be an expectation domain whose elements are defined as follows.  $U$  is the set of utility and disutility values  $\{U, u, -u, -U\}$  such that  $-U < -u < u < U$ .  $P$  is the set of plausibility values  $\{p, P\}$  such that  $p < P$ .  $\otimes$  is the multiplication-like operation that maps  $U \times P$  to  $T$ .  $T$  is therefore the set of product values  $\{-UP, -uP, -Up, -up, up, uP, Up, UP\}$  to be used in calculating plausibilistic expectation. These values are ordered according to the following specifications, where “|” is used analogously to its prior function in expressing incomparable plausibilities, utilities, and plausibilistic expectations:

1. Positive values:  $up < uP < UP; up < Up < UP; uP | Up$
2. Negative values:  $-UP < -uP < -up; -UP < -Up < -up; -uP | -Up$
3. Mixed values: for all  $x, y \in T$ ,  $(x < 0 \wedge y > 0) \rightarrow (x < y)$ .

$\oplus$  is the addition-like operation that maps  $T \times T$  to  $V$ . This operation, which is commutative, is defined for all  $x, y \in T$  and their absolute values  $|x|, |y|$  as follows:

1.  $(x < 0 \wedge y < 0) \rightarrow ((x \oplus y) = -)$ .
2.  $(x < 0 \wedge y > 0 \wedge |x| < |y|) \rightarrow ((x \oplus y) = +)$ .

3.  $(x < 0 \wedge y > 0 \wedge |x| = |y|) \rightarrow ((x \oplus y) = 0).$
4.  $(x < 0 \wedge y > 0 \wedge |x| > |y|) \rightarrow ((x \oplus y) = -).$
5.  $(x > 0 \wedge y > 0) \rightarrow ((x \oplus y) = +).$

The operation remains undefined for the sums “ $Up \oplus -uP$ ” and “ $-Up \oplus uP$ ” since the absolute values of the addends are incomparable. Finally,  $V$  is the set of plausibilistic expectation values  $\{-, 0, +\}$  ordered in the obvious way so that  $- < 0 < +$ .

Relative to  $D$ , we assume attributable states  $s_1, s_2, \dots, s_n$  and a plausibility function  $\pi$  that maps these states onto the values of  $P$ . In addition, we assume outcomes  $o_1, o_2, \dots, o_n$  and a utility function  $v$  that maps these outcomes onto the values of  $U$ . Then, the plausibilistic expectation ( $PE$ ) of an act  $a$  given evidence  $e$  can be defined as follows:

$$PE_{a,e} = \bigoplus_{i=1}^n v(o_i) \otimes \pi(s_i, e).$$

The right-hand side of  $PE$  is typographically identical to that of  $GEU$ , but their meanings are quite different;  $PE$  is a highly specific instance of  $GEU$ . The decision rule associated with  $PE$  is to maximize plausibilistic expectation.

To introduce the second and third decision rules, we need to pay closer attention to the consequences of order. In the foregoing remarks on order, we assumed that the  $\preceq$  relation establishes a partial order on utilities, for example. This permits a utility to be neither less than, equal to, or greater than another. Such utilities are therefore incomparable. The upshot is that, in such cases,  $PE$  cannot be applied. Let  $?_1$  and  $?_2$  be incomparable utilities and  $p_1$  and  $p_2$  comparable plausibilities. Then, the products in

$$(?_1 \otimes p_1) \oplus (?_2 \otimes p_2)$$

are incomparable, and the summation cannot be carried out (cf. Weirich 2004, p. 59). Like applications of  $E$  for expected utility, applications of  $PE$  require that outcome utilities be comparable. For analogous reasons,  $E$  demands comparable state probabilities and  $PE$  comparable state plausibilities. But utilities, probabilities, and plausibilities are sometimes incomparable in real-life decision making.

How might we deal with these incomparabilities in a rational way? One possibility is to shelve decision theory in any situation with incomparable plausibilities or utilities. A second possibility is to adapt the decision rule based on  $PE$  to the situation. Unfortunately, adopting the first option requires an answer to the question “What decision rule should be used instead?” Since I do not have a viable answer to this question, I favor the second option: adapt the decision rule to the situation.

There are actually two types of situation that call for adaptation. The first is where utilities are comparable while plausibilities are not; the second, where plausibilities are comparable but utilities are not. I will refer to the first type of situation as “utility-comparable” and to the second as “plausibility-comparable.” In both types of situation, I suggest, we should be guided by a common-sense notion: where just two criteria are relevant to a decision but one is inapplicable, we rely on the other criterion instead. Adherence to this notion in making real-life decisions would require ignoring any incomparable values, since nothing useful can be obtained from them, and relying on the comparable values.

In utility-comparable situations, the expectation domain  $D$  would become  $D_u = (U, P, T, V, \oplus, \otimes)$  with elements defined as follows.  $U$ ,  $V$ , and  $\oplus$  have the same meanings as for  $PE$ .  $P$  is the set of incomparable plausibility values  $\{p_1, p_2, \dots, p_n\}$ .  $\otimes$  is defined so that incomparable

plausibilities become right-identity elements; that is, for all  $u \in \mathbf{U}$  and all  $p \in \mathbf{P}$ ,  $u \otimes p = u$ . As a consequence,  $\mathbf{T} = \mathbf{U}$ .  $PE$  then reduces to the following definition of utility-comparable expectation ( $UCE$ ):

$$UCE_{a,e} = \bigoplus_{i=1}^n v(o_i).$$

The corresponding decision rule is to maximize utility-comparable expectation.

In plausibility-comparable situations, the expectation domain  $\mathbf{D}$  would become  $\mathbf{D}_p = (\mathbf{U}, \mathbf{P}, \mathbf{T}, \mathbf{V}, \bigoplus, \otimes)$ , defined analogously to  $\mathbf{D}_u$ .  $\mathbf{P}, \mathbf{V}$ , and  $\bigoplus$  retain their original meanings.  $\mathbf{U}$  is the set of incomparable utilities and disutilities  $\{u_1, u_2, \dots, u_n, -u_1, -u_2, \dots, -u_n\}$ .  $\otimes$  is defined so that incomparable utilities are left-identity elements and incomparable disutilities are negative left-identity elements; that is, for all  $u, -u \in \mathbf{U}$  and all  $p \in \mathbf{P}$ ,  $u \otimes p = p$  and  $-u \otimes p = -p$ . Therefore,  $\mathbf{T} = \{-P, -p, p, P\}$ , ordered in the obvious way. Accordingly,  $PE$  contracts to this definition of plausibility-comparable expectation ( $PCE$ ):

$$PCE_{a,e} = \bigoplus_{i=1}^n \pi(s_i, e).$$

The associated decision rule is to maximize plausibility-comparable expectation.

What I am proposing, then, is a suite of three decision rules: for fully comparable situations, maximize  $PE$ ; for utility-comparable situations, maximize  $UCE$ ; and for plausibility-comparable situations, maximize  $PCE$ . All three senses of expectation are special cases of Chu and Halpern's  $GEU$ . All three rules are motivated by the common-sense injunction "Use comparable data!"

*Relative disutility:* Jaakko Hintikka and Juhani Pietarinen propose to treat the epistemic utility of a hypothesis  $h$  and its contradictory  $\neg h$  as follows: "If  $h$  is true, the utility of his [the agent's] decision is the valid information he has gained. ... If  $h$  is false, it is natural to say that his disutility or loss is measured by the information he lost because of his wrong choice between  $h$  and  $\neg h$ , i.e., by the information he would have gained if he had accepted  $\neg h$  instead of  $h$ " (Hintikka and Pietarinen 1966, pp. 107–108; cf. Hintikka 1970, p. 16). This proposal for the utilities  $u$  of  $h$  and  $\neg h$  can be summed up by  $\circledast$  Table 21.1, where  $s_h$  and  $s_{\neg h}$  are states of the world posited by  $h$  and  $\neg h$ , respectively.

We are not concerned here with epistemic utility, but the Hintikka–Pietarinen proposal can be extended in the direction of our present interests. Because the choice of a hypothesis is an act, the proposal might be generalized for acts  $a_1$  and  $a_2$ , states  $s_1$  and  $s_2$ , and the outcome  $o_1$  of choosing  $a_1$  when  $s_1$  obtains and the outcome  $o_2$  of choosing  $a_2$  when  $s_2$  obtains. The generalized proposal can be summarized by  $\circledast$  Table 21.2.

We gain perspective on this generalization by noting the distinction between intrinsic utility and relative utility. (An analogous distinction can be drawn between intrinsic and

**Table 21.1**  
The Hintikka–Pietarinen proposal

	$s_h$	$s_{\neg h}$
$h$	$u(h)$	$-u(\neg h)$
$\neg h$	$-u(h)$	$u(\neg h)$

**Table 21.2****The Hintikka–Pietarinen proposal generalized**

	$s_1$	$s_2$
$a_1$	$u(o_1)$	$-u(o_2)$
$a_2$	$-u(o_1)$	$u(o_2)$

relative probability [Levi 1967, p. 98]). The intrinsic utility of some outcome is its utility considered in itself, without reference to other utilities. By contrast, the relative utility of an outcome is its utility compared to another utility. The intrinsic utility of an information outcome of 2 kb, say, would typically be small and positive because, considered in itself, it would be valued slightly. But the relative utility of the same outcome would be large and negative if we have to forego 1,000 kb of information in order to obtain it; that is, we would not prefer 2 kb of information but would greatly prefer 1,000. Hintikka and Pietarinen operate along the same relative lines. To say that the disutility of choosing  $h$  given  $s_h$  is  $-u(-h)$  and the disutility of choosing  $-h$  given  $s_h$  is  $-u(h)$  is to rely on relative disutilities.

Further perspective can be gained by underlining the distinction between the Hintikka–Pietarinen proposal and its foregoing generalization. Though the generalization is inspired by the Hintikka–Pietarinen proposal, there are several crucial differences. Here I mention only three.

The first is that Hintikka and Pietarinen's application of decision theory assumes numeric values for probabilities and utilities; by contrast, the decision theory outlined in these pages is comparative. It relies on comparative relations of plausibility, utility, and plausibilistic expectation. The practical import of this difference would be difficult to overstate. Because plausibilities and utilities can be specified using a bare minimum of comparative values, comparative decision theory can be applied in real-life situations that are just too amorphous for numeric forms of decision theory. It therefore signals an advance in the quest for a more realistic theory of decision recapped in the section  [Introduction](#).

The scope of the respective proposals constitutes a second difference. Hintikka and Pietarinen apparently meant for relative disutilities to be used in any choice between contradictory hypotheses. What I am proposing, by contrast, is spot duty. Relative disutilities can be useful in situations that are repeated over and over again in real-life decision making, situations in which utilities are only comparable in nonnumeric terms. That is, we may estimate one utility to be greater than another without being able to estimate how much greater it is. Relative disutilities are meant to be used in these situations. If, on the other hand, we can compare utilities in numeric terms, it goes without saying that we should use the more precise numeric comparison instead.

A third difference is that comparative decision theory generalizes the Hintikka–Pietarinen proposal in a sense we have yet to notice. Hintikka and Pietarinen were concerned with the binary case of contradictory hypotheses, but many options are not so neatly related. The propositions associated with nonidentical acts are often contraries instead of contradictories. Hence the comparative decision theory below generalizes Hintikka and Pietarinen's approach to cover more typical cases of choice where the propositions in play may be contraries as well as contradictories. This generalization is carried out as follows.

We begin by distinguishing total outcome, shared outcome, and unique outcome. An act's *total outcome* is its full set of consequences. An act's *shared outcome* is any part of its total

outcome that can be obtained by performing another act under consideration. An act's *unique outcome* is any part of its total outcome that cannot be obtained by performing other acts under consideration. Here is a simple example: if one act results in a share of stock and a samovar while another results in the same share of stock and a stone, the stock is the shared outcome of both acts; the samovar is the unique outcome of the first act; and the stone is the unique outcome of the second.

To expand the notion of relative disutility to options that may be contraries as well as contradictories, we note that the outcomes of acts  $a_1$  and  $a_2$  either overlap or not. If they do not, the total outcome of an act is identical to its unique outcome. This case would differ only slightly from the contradictory options of the Hintikka–Pietarinen scenario. That is, the outcome of choosing  $a_1$  would have any utility  $u_1$  produced by that act or the loss of any utility  $u_2$  provided by  $a_2$ . Alternatively, the outcome of choosing  $a_2$  would have utility  $u_2$  or disutility  $-u_1$ . Unlike the Hintikka–Pietarinen scenario, however, the propositions identified with  $a_1$  and  $a_2$  may be contraries, not contradictories, and so not jointly exhaustive. There may be a third option  $a_3$ . If so, we proceed as follows. Say that comparative decision theory permits the determination that  $a_1$  is inferior to  $a_2$ . The procedure can then be repeated for  $a_2$  and  $a_3$ . Like the outcomes of  $a_1$  and  $a_2$ , the outcomes of  $a_2$  and  $a_3$  either overlap or not. If they do not, we proceed as in this paragraph; if they do, we proceed as in the next one.

If the outcomes do overlap, the overlap may be total or partial. If total, there is no unique outcome because the acts lead to exactly the same outcome. Regardless of which act is chosen, then, the utility or disutility of this outcome would be the same:  $u$  or  $-u$ . If the overlap is partial, on the other hand, there are shared and unique outcomes. Because the shared outcome would be obtained whether the agent chooses  $a_1$  or  $a_2$ , it could not motivate the choice of one act rather than the other. According to the principle of independence, we would disregard the shared outcome and focus on unique outcomes. Suppose that  $a_1$  and  $a_2$  are acts whose associated propositions are contraries and that a unique outcome of  $a_1$  has utility  $u_1$  and a unique outcome of  $a_2$  has utility  $u_2$ . If the agent chooses  $a_1$ , she loses any utility offered by  $a_2$  but not by  $a_1$ ; she could have obtained this utility by choosing  $a_2$  instead. The disutility of this choice would be  $-u_2$ . Alternatively, if the agent chooses  $a_2$ , the disutility of this choice would be the loss of any utility uniquely provided by  $a_1$ . This disutility is  $-u_1$ .

In summary, even if the propositions associated with  $a_1$  and  $a_2$  are contraries, the outcomes of  $a_1$  and  $a_2$  are either completely distinct or not. If they are completely distinct, the situation is fundamentally no different for contraries than for contradictories. If the outcomes are not completely distinct, the overlap is total or partial. If total, the outcomes have the same utility or disutility; but if partial, the principle of independence authorizes choice based on unique outcomes alone. In cases of partial overlap, choice of one act relinquishes any unique outcome that would result from choice of the other. Hence any utility of an unattained unique outcome would also be lost. If the unique outcome of choosing one act has utility  $u$ , the relative disutility of the outcome of choosing the alternative act is  $-u$ . We can therefore generalize the Hintikka–Pietarinen proposal to include options that are contraries as well as contradictories.

## The Binary Case

Suppose that we are faced with a binary choice: act  $a_1$  or act  $a_2$ . Some of the states relevant to this choice may invite  $a_1$  in the sense that performing the act when these states obtain would

**Table 21.3****Binary cases for comparative decision theory**

Case	Plausibility	Utility
1	<	<
2	<	>
3	<	=
4	<	
5	>	<
6	>	>
7	>	=
8	>	
9	=	<
10	=	>
11	=	=
12	=	
13		<
14		>
15		=
16		

produce desirable outcomes – desirable relative to other possible outcomes. Other states may invite  $a_2$  in an analogous sense. For example, if you are debating whether to take your umbrella or not, the state of rain today invites taking the umbrella and the state of no rain today invites leaving it at home. The plausibilities of these relevant states exhibit relations of infraplausibility ( $<$ ), supraplausibility ( $>$ ), equiplausibility ( $=$ ), or incomparability ( $|$ ). Structurally analogous relations hold among the utilities of the outcomes given the various act-state pairs. Hence there are 16 possible cases.

These cases are summarized in [Table 21.3](#). In order to highlight essentials, the table employs abbreviations. For example, “ $<$ ” in the plausibility column abbreviates “ $\pi(s_1, e) < \pi(s_2, e)$ ,” which says that the plausibility of the  $s_1$  states that invite  $a_1$  given the total evidence  $e$  is less than the plausibility of the  $s_2$  states that invite  $a_2$  given  $e$ . In other words, plausibility considerations favor  $a_2$ . Similarly, “ $<$ ” in the utility column is short for “ $v(o_1) < v(o_2)$ ,” which says that the utility of outcome  $o_1$  from choosing  $a_1$  is less than the utility of outcome  $o_2$  from choosing  $a_2$ . In this case, utility considerations favor  $a_2$ .

These cases fall naturally into seven groups, the first of which is formed by cases 1 and 6. Let “ $U$ ” and “ $u$ ” express higher and lower utility while “ $P$ ” and “ $p$ ” stand for higher and lower plausibility. In case 1,  $a_1$  provides utility  $u$  with plausibility  $p$  and disutility  $-U$  with plausibility  $P$ , but  $a_2$  yields utility  $U$  with plausibility  $P$  and disutility  $-u$  with plausibility  $p$ . To choose between the acts, we employ the decision rule of maximizing plausibilistic expectation. Where plausibilistic expectation is defined by  $PE$ , the plausibilistic expectation of  $a_1$  is

$$PE_1 = up \oplus -UP = -.$$

Similarly,  $PE$  determines the plausibilistic expectation of  $a_2$ :

$$PE_2 = -up \oplus UP = +.$$

Because the plausibilistic expectation of  $a_1$  is negative while that of  $a_2$  is positive,  $a_1$  is inferior to  $a_2$ . A parallel argument for case 6 shows that  $a_2$  is inferior to  $a_1$ .

The second group includes cases 3, 7, 9, and 10, which have one relation that is  $=$  while the other is either  $<$  or  $>$ . Suppose that there are two routes to a vacation destination and that the only relevant criteria for choice are cost and comfort. Suppose also that the routes are equally costly but one is more comfortable than the other. Then the more comfortable route should be chosen. Generally, in any choice with two relevant criteria and a tie with respect to one of them, common sense recommends that the other criterion be decisive. Case 3 follows this advice. Here  $a_1$  offers utility  $U$  with plausibility  $p$  and disutility  $-U$  with plausibility  $P$ ; for  $a_2$ , the plausibilities are reversed. Because the acts are tied with respect to utility, the tie is broken by plausibility. According to  $PE$ , the plausibilistic expectation of  $a_1$  is

$$PE_1 = Up \oplus -UP = -,$$

while that of  $a_2$  is

$$PE_2 = -Up \oplus UP = +.$$

Since the plausibilistic expectation of  $a_2$  is positive and that of  $a_1$  is negative,  $a_2$  is superior to  $a_1$ .

The sparsely populated third group is made up of case 11 alone. This case differs from those of the second group because not just one but both relations are  $=$ .  $PE$  would give the plausibilistic expectation of both acts as

$$PE = UP \oplus -UP = 0.$$

Note that an expectation of 0 does not imply that the associated act has no intrinsic utility or disutility. Like our other expectations, this expectation is comparative. Hence there is no comparative advantage for either act – a tie.

Cases 13, 14, and 15 comprise a fourth group, whose members have a plausibility relation that is  $\mid$  and a utility relation that is not  $\mid$ . Since these cases are utility-comparable, the relevant sense of expectation is  $UCE$ . We thereby ignore the incomparable plausibilities and base the decision on utility alone. In case 13, for example, the utility-comparable expectation of  $a_1$  would be

$$UCE_1 = u \oplus -U = -,$$

while that of  $a_2$  would be

$$UCE_2 = -u \oplus U = +.$$

Because the utility-comparable expectation of  $a_1$  is negative while that of  $a_2$  is positive,  $a_2$  is the better choice.

A fifth group contains cases 4, 8, and 12, which have a utility relation that is  $\mid$  and a plausibility relation that is not  $\mid$ . Because the members of this group are plausibility-comparable, the appropriate definition of expectation is  $PCE$ . Applying it to case 4 gives the plausibility-comparable expectation of  $a_1$  as

$$PCE_1 = p \oplus -P = -$$

and that of  $a_2$  as

$$PCE_2 = -p \oplus P = +.$$

Since the plausibility-comparable expectation of  $a_2$  is positive and that of  $a_1$  is negative, the choice should be  $a_2$ .

Cases 2 and 5 constitute a sixth group. Its members are heterogeneous in that each act compares favorably in one respect but unfavorably in the other. Consider case 2, for instance. Where  $a_1$  has more utility but less plausibility, its plausibilistic expectation would be

$$PE_1 = Up \oplus -uP.$$

By contrast, the plausibilistic expectation of  $a_2$  would be

$$PE_2 = -Up \oplus uP.$$

As noted in the [Preliminaries](#) subsection of [Current Research](#), the  $\oplus$  operation is undefined in both cases. There is no way to compare the plausibilistic expectations without knowing how much more desirable the outcomes of  $a_1$  are and how much more plausible the states that invite  $a_2$  are. In these cases, therefore, we can reach no decision.

Only case 16 remains. Like cases 2 and 5, it results in no decision, but it does so for a different reason. In cases 2 and 5, comparative decision theory breaks down. In case 16, however, it cannot even start up; since both plausibility and utility are incomparable, comparative decision theory – like any other form of decision theory – can say nothing at all.

These three problem cases are not equally problematic. Unlike case 16, cases 2 and 5 can sometimes be resolved on a case-by-case basis. Take a decision about whether or not to attend an academic conference, for example. Although serious accidents can happen at home, venturing out onto the highways and airways of the conference circuit most likely increases the plausibility of a serious accident. Hence plausibilistic considerations would encourage us to stay home, yet many of us decide not to do so. Why? Because we take the far greater utility of a conference to outweigh the slightly greater plausibility of a travel accident. On the other hand, if we were to judge the plausibility of a travel accident to be significantly higher than the plausibility of an accident at home, we would choose to stay home. This is a rational choice provided we assign a normal (very great) disutility to serious bodily harm. The writer once opted out of a conference in Turkey because he estimated the plausibility of war spilling over from northern Iraq to Turkey to be decidedly higher than the plausibility of an accident at home.

Note that our results acknowledge the difference between indifference and indecision. Cases 11, 12, and 15 result in indifference: there is a good decision-theoretic reason to choose  $a_1$  and a good decision-theoretic reason to choose  $a_2$ . But cases 2, 5, and 16 terminate in indecision: there is no decision-theoretic reason to choose  $a_1$  and no decision-theoretic reason to choose  $a_2$ .

It is sometimes overlooked that even though the indifference judgments in cases 11, 12, and 15 do not assert that  $a_1$  is superior to  $a_2$  or vice versa, they do make an assertion. They assert that  $a_1$  is as good as  $a_2$ . This is a disjunctive judgment, analogous to the disjunctive solutions proposed in the literature on moral dilemmas (Greenspan 1983, pp. 117–118; Gowans 1987, p. 19; Zimmerman 1996, pp. 209, 220–221). Disjunctive judgments do valuable cognitive work. Consider the four basic possibilities for binary choice of any kind: option 1, option 2, both option 1

■ **Table 21.4**  
Binary cases with resolutions

Case	Plausibility	Utility	Resolution
1	<	<	$a_2$
2	<	>	No decision
3	<	=	$a_2$
4	<		$a_2$
5	>	<	No decision
6	>	>	$a_1$
7	>	=	$a_1$
8	>		$a_1$
9	=	<	$a_2$
10	=	>	$a_1$
11	=	=	$a_1$ or $a_2$
12	=		$a_1$ or $a_2$
13		<	$a_2$
14		>	$a_1$
15		=	$a_1$ or $a_2$
16			No decision

and option 2, neither option 1 nor option 2. An agent who forms the judgment “ $a_1$  or  $a_2$ ,” as in cases 11, 12, and 15, has already rejected the neither option. And, since the acts cannot be chosen simultaneously, the agent *might* feel compelled by circumstances to choose one of them even though she has no reason to choose it over the alternative. The agent would have made a disjunctive judgment that excludes two of the four basic options: neither and both.

The results of our discussion can now be summarized in ▶ **Table 21.4**. Of the 16 cases represented there, 13 are resolvable in comparative terms; only 3 are not.

## The Finite General Case

One of the assumptions in the ▶ **Preliminaries** subsection is that real-life decision makers can consider only a finite number of acts at a given moment. This makes the foregoing analysis of the binary case critical. For if it is possible to choose between  $a_1$  and  $a_2$  such that  $a_2$ , say, is the winner, then it is also possible in principle to choose between  $a_2$  and any act  $a_3$ . The winner can then be compared with any act  $a_4$ , and so on.

Here we rely on the additional assumption that act preference is transitive. The transitivity of preference is routinely affirmed by decision theorists (Savage 1972, p. 18; Jeffrey 1983, pp. 144–145; Maher 1993, p. 60), yet this affirmation has been repeatedly challenged (e.g., Hughes 1980; Black 1985; Baumann 2005). Since transitivity is discussed in greater depth in the section on ▶ **Further Research**, for the moment I will limit myself to the two following claims. First, even if act preference should turn out to be intransitive, binary choice could still be

completed on comparative grounds, for transitivity is irrelevant when there are only two options. Comparative decision theory can always be applied to any two acts under consideration. Second, the assumption that act preference is transitive, if properly understood and suitably employed, does in fact hold. The main consideration is to restrict transitive inference to the same sense of “preference.” That is, we need to avoid equivocation.

To see the damage that equivocation can wreak, let us consider a relatively transparent instance. Max Black attempted to show that intransitive preferences can be rational by instancing job candidates A, B, and C who are rated for expertise, congeniality, and intelligence on a scale of 1–3, where 3 is high (1985). Their scores for each characteristic in the order mentioned are as follows:

A: 3, 2, 1

B: 1, 3, 2

C: 2, 1, 3.

Given these scores, an employer would prefer A to B (for expertise), B to C (for congeniality), and C to A (for intelligence). Hence, it appears, transitivity is violated but not rationality.

That transitivity is violated is a mere appearance, however, thrown up by simple equivocation. We have cycled from preferred-for-expertise to preferred-for-congeniality to preferred-for-intelligence. Jumbling these three senses of preference together can create problems in much the same way as mixing binary, decimal, and hexadecimal numerals. Decision-theoretic preference is not preference for one characteristic and then another; it is preference overall. “I am concerned with preference *all things considered*, so that one can prefer buying a Datsun to buying a Porsche even though one prefers the Porsche *qua* fast (e.g., since one prefers the Datsun *qua* cheap, and takes that desideratum to outweigh speed under the circumstances)” (Jeffrey 1983, p. 225). If an employer were to have an *overall* preference for A to B, B to C, and C to A, we would have a genuine violation of transitivity. But we would also have a violation of rationality.

The moral: equivocation can produce apparent violations of transitivity. This issue is treated in more detail in the section on [Further Research](#).

## Further Research

The comparative decision theory just outlined is based on the assumptions identified in the [Preliminaries](#) subsection of [Current Research](#). However, most of these assumptions were simply postulated without any attempt at justification. Though this cannot be remedied completely here, I do want to inquire into three of these assumptions in particular: transitivity, independence, and the proposed decision rules. These assumptions mark choice nodes even for those working within alternative decision-theoretic frameworks. Hence the following remarks are offered as suggestions for further research.

## Transitivity

Within the confines of comparative decision theory, transitivity can be an issue at three different levels. At the most basic level, if plausibility  $p_1$  (or utility  $u_1$ ) is greater than plausibility

$p_2$  (or utility  $u_2$ ), and  $p_2$  (or  $u_2$ ) is greater than plausibility  $p_3$  (or utility  $u_3$ ), then we might infer that  $p_1$  (or  $u_1$ ) is greater than  $p_3$  (or  $u_3$ ). Such inferences, including analogous inferences with the relations of equal to, less than, and incomparable to, exhibit *factor transitivity*. At a second level, whenever a product  $r_1$  of utility and plausibility is equal to another such product  $r_2$  and  $r_2$  is equal to a third such product  $r_3$ , we may conclude that  $r_1$  is equal to  $r_3$ . Along with similar inferences involving the relations greater than, less than, and incomparable to, these inferences instantiate *product transitivity*. Finally, if the plausibilistic expectation of act  $a_1$  is less than the plausibilistic expectation of act  $a_2$  and that of  $a_2$  is less than that of act  $a_3$ , then we may deduce that the plausibilistic expectation of  $a_1$  is less than that of  $a_3$ . Such inferences, including parallel inferences with the relations of greater than, equal to, and incomparable to, display *expectation transitivity*.

Factor transitivity and product transitivity are both relatively transparent and relatively peripheral to our present concerns. They are relatively transparent because the inferences involve a single, clearly articulated relation:  $>$ ,  $<$ ,  $=$ , or  $\mid$  as already defined for factor relata (in the [Preliminaries](#) subsection of [Current Research](#)) and analogously definable for product relata. For example, if one plausibility is greater than a second and this second is greater than a third, the plausibility of the first must be greater than the third. Nevertheless, factor transitivity and product transitivity are relatively peripheral for our purposes because the extension of comparative decision theory from the binary to the finite general case appeals to expectation transitivity, not factor or product transitivity. We want to be able to infer, say, that the expectation of act  $a_1$  is equal to that of act  $a_3$  simply because the expectation of  $a_1$  is equal to that of act  $a_2$  and that of  $a_2$  is equal to that of  $a_3$ . The inference permits a comparative evaluation of  $a_1$  and  $a_3$  without having to compare them directly.

Let us therefore turn to expectation transitivity. Our discussion requires attention to three types of *homogeneous groups*. *Fully comparable groups* are composed of one or more binary comparisons in which the utilities and plausibilities are comparable in terms of greater than, equal to, or less than. For example, a comparison of act  $a_1$  and act  $a_2$  where all utilities are comparable and all plausibilities are comparable would constitute a fully comparable group. *Utility-comparable groups* are formed by one or more binary comparisons in which the utilities are comparable while the plausibilities are incomparable. *Plausibility-comparable groups* consist of one or more binary comparisons in which the plausibilities are comparable but the utilities are incomparable.

Fully comparable groups appear to pose no problems for transitivity. The relata, which are plausibilistic expectations ( $PE$ ), form a homogeneous set. If we are considering acts  $a_1$ ,  $a_2$ , and  $a_3$  where  $PE(a_1) < PE(a_2)$  and  $PE(a_2) < PE(a_3)$ , then  $PE(a_1) < PE(a_3)$ .

The other homogeneous groups are similarly transparent. Consider acts  $a_1$ ,  $a_2$ , and  $a_3$  with outcome utilities that are comparable but state plausibilities that are incomparable. Since this is a utility-comparable group, the appropriate decision rule is based on utility-comparable expectation ( $UCE$ ). Let  $UCE(a_1) > UCE(a_2)$  and  $UCE(a_2) > UCE(a_3)$ . Then, straightforwardly,  $UCE(a_1) > UCE(a_3)$ . Parallel remarks apply to plausibility-comparable groups.

Unfortunately, not all groups are as well behaved. The decision rules based on  $PE$ ,  $UCE$ , and  $PCE$  all assume homogeneity: for a given choice, all the utilities and plausibilities are fully comparable, or they are all utility-comparable, or they are all plausibility-comparable. At times, however, some utilities and plausibilities may be fully comparable while others can be either utility-comparable or plausibility-comparable. These heterogeneous decision problems are characterized by *mixed groups*.

**Table 21.5** **$a_1$  or  $a_3$ ?**

	$\pi(s_1) = p$	$\pi(s_2) = P$
$a_1$	$v(o_1) = U$	$v(o_2) = -u$
$a_3$	$v(o_3) = -U$	$v(o_4) = u$

Watch what can happen when we attempt transitive inference across mixed groups. Let acts  $a_1$  and  $a_2$  be utility-comparable such that  $UCE(a_1) > UCE(a_2)$ . In addition, acts  $a_2$  and  $a_3$  are utility-comparable such that  $UCE(a_2) > UCE(a_3)$ . But let  $a_1$  and  $a_3$  be fully comparable. Their comparison is represented by [Table 21.5](#), where  $\pi$  is a plausibility function;  $s_1$  and  $s_2$  are relevant states;  $p$  and  $P$  are state plausibilities such that  $p < P$ ;  $v$  is a utility function;  $o_1, o_2, o_3$ , and  $o_4$  are outcomes of act-state pairs; and  $U, u, -u, -U$  are outcome utilities such that  $-U < -u < u < U$ . According to the analysis in the section [Current Research](#), the comparison of  $a_1$  and  $a_3$  should result in no decision because the decision-theoretic verdict is split: utility considerations favor  $a_1$ , but plausibility considerations favor  $a_3$  (cf. case 2 of [Table 21.4](#)). But, having noted that the expectation of  $a_1$  is greater than that of  $a_2$  and that of  $a_2$  is greater than that of  $a_3$ , we might venture the transitive inference that the expectation of  $a_1$  is greater than that of  $a_3$ . This would be inconsistent with the conclusion that the pairwise comparison of  $a_1$  and  $a_3$  results in no decision.

What has gone wrong? The answer, in a word, is equivocation. The transitive inference that the expectation of  $a_1$  is greater than that of  $a_3$  is fallacious. It equivocates by conflating the utility-comparable expectations of  $a_1$  relative to  $a_2$  and  $a_2$  relative to  $a_3$  with the fully comparable expectation of  $a_1$  relative to  $a_3$ . Instead, we should conclude that even though  $a_1$  is decision-theoretically superior to  $a_2$  and  $a_2$  is decision-theoretically superior to  $a_3$ , there is no decision-theoretic reason to prefer  $a_1$  over  $a_3$  or vice versa – unless, of course, we are willing to place more weight on either plausibility or utility.

Clearly, then, we need to restrict expectation transitivity. The restriction is that expectation transitivity must be limited to homogeneous groups. Transitivity can be invoked if all the comparisons are fully comparable, or if they are all utility-comparable, or if they are all plausibility-comparable. In insisting on this restriction, we are merely insisting on the same sense of “expectation” in each case. Fully comparable expectation, utility-comparable expectation, and plausibility-comparable expectation are different, though closely related, concepts. Mixing them up can generate fallacies.

The foregoing remarks make no claim to have proved the transitivity of preference; the issue is far too complex for that. But the transitivity assumption seems to be as widely accepted as any normative principle of rational choice; it is common to both the Anglo-American and Franco-European schools of decision theory, for instance (Fishburn 1991, p. 115). Nevertheless, this is a topic for further research.

## Independence

Another main assumption of the decision theory outlined in the section on [Current Research](#) is the principle of independence. This principle, which licenses ignoring shared

outcomes and concentrating on unique outcomes, is controversial. It figured as one of Savage's postulates under the guise of "the sure-thing principle" (1972, p. 21), was targeted by the Allais and Ellsberg paradoxes (Allais 1953, 1979a, b; Ellsberg 1961), and continues to be affirmed in one form or another by Jeffrey (1983, p. 23), Levi (1986, pp. 129, 144), and Maher (1993, pp. 12, 83). While a full-blown investigation would be out of place, I do want to acknowledge the controversy and say just a few words about it. The discussion proceeds through three stages: a brief exploration of the Allais and Ellsberg paradoxes; a search for solid ground for the principle of independence; and a suggestion about porting trickier cases to this solid ground.

Some adherents of the principle of independence objected to Allais' original counter-examples because they unrealistically require ordinary people to make hypothetical choices with potential payoffs of hundreds of millions of dollars (e.g., Morgenstern 1979, pp. 178, 180; Amihud 1979, pp. 151–152). In response, Allais pointed out that maximizing expected utility can also create problems in realistic situations. The paradox introduced in the following paragraph is an instance. Even though it does not rely directly on independence, I concentrate on it because of its simplicity and proximity to ordinary reasoning. The treatment of this simple example can be extended in obvious ways to more complicated scenarios that do rely directly on independence, but there is no need to carry out this extension here.

Allais claims that "a person who is not generally considered as irrational, faced with a single, non-renewable choice, may well take ten dollars in cash rather than gamble on an even chance of winning \$22 or nothing" (1979b, p. 539). Such a person, that is, might choose the sure \$10 despite the fact that the choice is not the expected utility winner, and such a choice need not be irrational. In this and other examples, Allais' point is that maximizing expected utility neglects "the impact of the greater or lesser propensity for risk-taking or security, the consequence of which is, in particular, a complementarity effect in the neighbourhood of certainty" (1979b, p. 442). In short, the maximizing approach ignores "the considerable psychological importance attaching to the advantage of certainty as such" ([1953]1979a, p. 88).

The first thing to be noticed about this example is that it trades on a mistake: the conflation of monetary outcomes and utilities. In order to make the point that the choice of \$10 is not the expected utility winner, Allais assumes outright that the utilities of the outcomes of \$0, \$10, and \$22 are 0, 10, and 22, respectively. At least since the time of Daniel Bernoulli, however, economists have recognized that the relation between money and its utility is not necessarily linear (cf. Levi 1986, p. 142). Hence the monetary outcomes of the example may or may not have utilities of 0, 10, and 22. For the sake of the argument, however, I will assume that they do.

The notion that acts should be evaluated by their consequences has been called a decision-theoretic "pre-axiom" by Peter Hammond (1988, p. 73), who argues that this consequentialist presupposition implies a number of standard axioms, including some forms of the principle of independence. Hence non-consequentialist preferences – preferences for a state or an act, for example – can lead to violations of these axioms. An example of a state-dependent preference is a favorable view of a state of financial crisis because the Joneses would be poorer and the task of keeping up with them less onerous (cf. Hirshleifer 1965, p. 532). Perhaps, then, the security motivating those who would choose the sure \$10 is a state-dependent preference.

However, as one writer on state-dependent preferences observed, "states may vary in respect to 'nonpecuniary income'" (Hirshleifer 1965, p. 532). This is a revealing observation. If income is a decision-theoretic consequence and there are nonpecuniary incomes, then such preferences may actually be consequence-dependent rather than state-dependent. Suppose we explore this idea in the context of the Allais paradoxes (cf. Maher 1993, p. 82). Consider our

**Table 21.6****An Allais problem with outcomes**

	$s_1$	$s_2$
Accept the \$10	\$10, FA	\$10, FA
Accept the gamble	\$22, FA	\$0, -FA

example with the sure \$10 from the point of view of someone for whom \$10 would be extremely important – someone who would urgently need the money for first aid, for example. In such a case, we would need to include the outcomes of first aid (FA) and no first aid ( $-FA$ ) in addition to the already-specified monetary outcomes. The agent's predicament could then be represented by [Table 21.6](#).

What would be the utilities of these four outcomes? If we adapt the comparative resources developed in the section on [Current Research](#), the answer is straightforward: For a utility function  $v$  with values  $U > u > -U$ ,

$$\begin{aligned} v(\$22, \text{FA}) &= U \\ v(\$10, \text{FA}) &= u \\ v(\$0, -\text{FA}) &= -U. \end{aligned}$$

The third utility assignment is based on the consideration that all utility from the outcome (\$22, FA) would be lost by accepting the gamble when  $s_2$  obtains. This is actually a variation on the theme of relative disutility introduced in the section on [Current Research](#). The original concept of relative disutility could be called *column disutility*: disutility relative to other values in the column that represent what the agent might have enjoyed if she had acted differently. By contrast, the relative disutility employed in the third utility assignment would be *row disutility*: disutility relative to other values in the row that represent what the agent might have obtained had the world been different. Note that since there is no overlap between the outcomes (\$22, FA) and (\$0,  $-FA$ ), there is no call to apply the principle of independence here.

Substituting these utilities for their associated outcomes in [Table 21.6](#) yields [Table 21.7](#).

We can now calculate the expected utilities  $E$  of the two acts. Since the probability of each state is  $\frac{1}{2}$ ,  $E(\text{accept the } \$10) = u$ , and  $E(\text{accept the gamble}) = 0$ . The security-conscious agent we are considering should therefore accept the \$10 (though other agents faced with other outcomes might reasonably accept the gamble). We have arrived at the secure Allais result, but we have done so by simple maximization of expected utility.

Let us turn briefly to the Ellsberg paradox (1961, pp. 653–656). Imagine an urn known to contain 30 red balls and 60 black and yellow ones; the proportion of black to yellow is unknown. One ball is to be drawn at random from the urn. You are faced with two choices. Choice 1 is between act  $a_1$ , to bet on red with payoffs of \$100 if you win and \$0 if you lose, and act  $a_2$ , to bet on black with payoffs of \$100 if you win and \$0 if you lose. Choice 1 is summarized by [Table 21.8](#). Choice 2 is between act  $a_3$ , to bet on red or yellow with payoffs of \$100 if you win and \$0 if you lose, and act  $a_4$ , to bet on black or yellow with payoffs of \$100 if you win and \$0 if you lose. Choice 2 is summarized by [Table 21.9](#).

Unlike the Allais problem we have just discussed, the Ellsberg problem directly challenges the principle of independence. According to it, the yellow column in both choices should be

**Table 21.7****An Allais problem with utilities**

	$s_1$	$s_2$
Accept the \$10	$u$	$u$
Accept the gamble	$U$	$-U$

**Table 21.8****The Ellsberg paradox: choice 1**

	Red	Black	Yellow
$a_1$ bet on red	\$100	\$0	\$0
$a_2$ bet on black	\$0	\$100	\$0

**Table 21.9****The Ellsberg paradox: choice 2**

	Red	Black	Yellow
$a_3$ bet on red or yellow	\$100	\$0	\$100
$a_4$ bet on black or yellow	\$0	\$100	\$100

ignored since its payoffs in Choice 1 are the same and its payoffs in Choice 2 are the same. If that is done, Choice 1 and Choice 2 become numerically identical. Hence anyone who prefers  $a_1$  over  $a_2$  should also prefer  $a_3$  over  $a_4$ . But Ellsberg reports that many people prefer  $a_1$  to  $a_2$  yet also prefer  $a_4$  to  $a_3$ . Others, though fewer, prefer  $a_2$  to  $a_1$  yet also prefer  $a_3$  to  $a_4$ . Both preference patterns violate the principle of independence.

I submit that the twist in the Ellsberg problem is that both Choice 1 and Choice 2 are instances of what I will call *laminated choice*. Each choice is constituted by a superposition of two layers, but one of these layers is explicit while the other is implicit. In Choice 1, the explicit layer is fully accounted for by [Table 21.8](#) above. But the implicit layer is a further choice.

This further choice hinges on the difference between definite probabilities, such as the probability of drawing a red ball in this situation, and indefinite probabilities, such as that of drawing a black ball. The acts under consideration are to bet with definite probabilities and to bet with indefinite probabilities. The possible outcomes for this choice are a better chance, a worse chance, and an unchanged chance to win the bet. States, as we have noted, can be thought of as propositions, and the states relevant to the choice at hand are states that affect the world's predictability. Of particular interest in this case are the states that definite probabilities facilitate accurate prediction more than indefinite probabilities, that indefinite probabilities facilitate accurate prediction more than definite probabilities, and that neither type of probability facilitates accurate prediction more than the other. For brevity, I will refer to these states of the world as "favors definite," "favors indefinite," and "favors neither." The choice in the implicit layer of the problem can then be summarized by [Table 21.10](#).

**Table 21.10****The Ellsberg paradox: choice 1's implicit layer**

	Favors definite	Favors indefinite	Favors neither
$a_5$ bet with definite probabilities	Better chance to win	Worse chance to win	Unchanged chance to win
$a_6$ bet with indefinite probabilities	Worse chance to win	Better chance to win	Unchanged chance to win

What I am suggesting, then, is that Choice 1 can be summarized by two decision tables:

➊ [Table 21.8](#), which features  $a_1$  and  $a_2$ , and ➋ [Table 21.10](#), which features  $a_5$  and  $a_6$ . An analogous point holds for Choice 2. These laminated choices are possible because acts are subject to multiple true descriptions. One act can be truly described as both  $a_1$ , a bet on red, and  $a_5$ , a bet with definite probabilities. The decision tables for choices defined in these alternative terms are, as it were, superposed.

The superposition of choices can be made more transparent by noting the distinction that Ellsberg builds his entire analysis around: Frank Knight's distinction between measurable uncertainty or risk, which can be expressed by numerical probabilities, and unmeasurable uncertainty, which cannot ([1921]1971, pp. 19–20). The statistical probabilities of 1/3 for a red ball and 2/3 for a black or yellow ball are measurable uncertainties, but Ellsberg thinks that the problem's residual uncertainty is not probabilistic and not measurable (1961, p. 659).

Ellsberg is right, I believe, to think that there are two types of uncertainty here, but I would develop the contrast differently. All probabilities are plausibilities, as we have noted, but not all plausibilities are probabilities. Hence there are probabilistic and nonprobabilistic plausibilities. Both are present in the Ellsberg problem: the numerical probability of states like the drawing of a red ball and the comparative plausibility of states like favoring bets made with definite probabilities. The explicit layer of the problem relies on numerical probabilities; the implicit layer, on comparative plausibilities.

Looked at in this way, the Ellsberg problem no longer appears to violate the principle of independence. Take Choice 1, understood as a superposition of the choice between  $a_1$  and  $a_2$  and the choice between  $a_5$  and  $a_6$ . The choice between  $a_1$  and  $a_2$  cannot be made by maximizing expected utility ( $E$ ). Although  $E$  of  $a_1$  could be determined, that of  $a_2$  could not, for  $E$  requires a definite probability for drawing a black ball, and that we do not have. Although it would be possible to estimate the probability of black in various ways – by defining upper and lower probability measures, for instance (Halpern 2003, pp. 25–28) – unless we are prepared to work with multiple probability measures and to recast our decision rule to accommodate them, there is no solution at the explicit layer.

But there is a solution at the implicit layer. Given that the utilities of the outcomes of  $a_5$  and  $a_6$  are evenly balanced, those who would prefer  $a_5$  can do so reasonably if and only if they hold a plausibility function  $\pi$  that returns these comparative plausibilities of states:

$$\pi(\text{favors definite}) > \pi(\text{favors indefinite}).$$

That is, they would choose  $a_5$  because they believe it offers them a better chance of winning the bet, and they believe this because, in effect, they are maximizing plausibilistic expectation ( $PE$ ). Given  $\pi$ ,  $a_5$  turns out to maximize  $PE$ . Analogously, Choice 2 cannot be made by maximizing  $E$ .

**Table 21.11****Grounds for the principle of independence**

	$s_1$	$s_2$	$s_3$
$a_1$	$o_1$	$o_2$	$o_3$
$a_2$	$o_4$	$o_5$	$o_3$

for the choice between  $a_3$  and  $a_4$ , but it can be made by maximizing  $PE$  for the choice between  $a_5$  and  $a_6$ . In both cases, those who opt for  $a_5$  could do so for the same plausibilistic reason. And they could do so without violating the principle of independence.

The second stage of our discussion of the principle of independence is a search for solid ground for the principle. To initiate this search, consider the usual decision-theoretic case where state probabilities are independent of acts. As a simple illustration, take a case that is structurally similar to the Ellsberg problem. The possible outcomes  $o_1$ – $o_5$  of acts  $a_1$  and  $a_2$  given states  $s_1$ ,  $s_2$ , and  $s_3$  are reflected in [Table 21.11](#), where  $o_1$  and  $o_4$ , on the one hand, and  $o_2$  and  $o_5$ , on the other, are assumed to be nonidentical. Since the outcomes in the  $s_3$  column are identical, their utilities are identical also. Provided that the probability of  $s_3$  does not vary with the choice of  $a_1$  or  $a_2$ , the products  $r_3$  formed by  $o_3$ 's utility and  $s_3$ 's probability must therefore be identical as well. Where the remaining products of utility and probability are expressed in the obvious way, the expected utility  $E$  of the two acts would be:

$$\begin{aligned} E(a_1) &= r_1 + r_2 + r_3 \\ E(a_2) &= r_4 + r_5 + r_3. \end{aligned}$$

Consequently, the relative magnitude of the acts' expected utilities is independent of the products  $r_3$  – just as the independence principle says. The same point holds for plausibilistic expectation. In these cases, then, independence is no more than elementary algebra. When state probabilities do not vary with acts, the principle of independence is on entirely solid ground (cf. Jeffrey 1983, p. 23).

The final stage of our discussion of independence concerns the remaining question: What if state probabilities do vary with acts? Here, of course, independence need not hold. Michael D. Resnik describes a decision about whether or not to smoke where the relevant states are contracting lung cancer and not contracting lung cancer (1987, pp. 15–16). Evidently, the probabilities of these states do vary with the acts of smoking and not smoking. But Resnik thinks the problem should be reformulated. Since not all smokers get lung cancer, there must be some protective factor that some people have and others do not. So Resnik proposes replacing the states of getting lung cancer and not getting lung cancer with four states related to this protective factor: having the protective factor and getting lung cancer from nonsmoking causes; having the protective factor and not getting lung cancer from nonsmoking causes; not having the protective factor and getting lung cancer from nonsmoking causes; and not having the protective factor and not getting lung cancer from nonsmoking causes (1987, p. 16). The probabilities of these states do not vary with the acts of smoking and not smoking.

I also think that the problem should be reformulated, but my suggestion is different. From the point of view of the person trying to decide whether to smoke, getting lung cancer and not getting lung cancer are not states at all. They are outcomes. The relevant states, on the other

hand, can be very roughly described as having a predisposition to lung cancer and not having a predisposition to lung cancer. The probabilities of these states, like the probabilities of Resnik's states, do not vary with the acts of smoking and not smoking. Consequently, when the decision is conceptualized in these terms, the principle of independence can be unproblematically applied.

In sum, the suggestion for dealing with states whose probabilities vary with acts is to attempt to reformulate them as states whose probabilities do not vary with acts. Whether this strategy can always be employed, or if not, when it can and cannot be employed, are questions for further research.

## Decision Rules

The suite of decision rules proposed in the [Preliminaries](#) subsection of [Current Research](#) keyed on plausibilistic expectation (*PE*), utility-comparable expectation (*UCE*), and plausibility-comparable expectation (*PCE*). As we have noted, all three rules assume homogeneity: for a given choice, all the utilities and plausibilities are fully comparable, or they are all utility-comparable, or they are all plausibility-comparable. But our discussion of transitivity has already adduced one instance of a mixed group. In mixed groups, some utilities and plausibilities may be fully comparable while others may be either utility-comparable or plausibility-comparable. Suppose, for example, that an agent is considering an act whose relevant states  $s_1$ ,  $s_2$ , and  $s_3$  have plausibilities  $p_1$ ,  $p_2$ , and  $p_3$ , respectively. Performing the act will result in one of the outcomes  $o_1$ ,  $o_2$ , and  $o_3$ , which have utilities  $u_1$ ,  $u_2$ , and  $u_3$ , respectively. Assume that  $p_1-p_3$  are comparable,  $u_1$  and  $u_2$  are comparable, but  $u_3$  is incomparable to  $u_1$  and  $u_2$ . What decision rule(s) should be used?

One possibility is to adopt the norm "For mixed groups, mix the rules." That is, use the rule based on *PE* whenever possible and the rules based on *UCE* and *PCE* whenever necessary. In the case just described, this would require calculating *PE* for the fully comparable possibilities with  $u_1$  and  $u_2$  but *PCE* for the plausibility-comparable possibility with  $u_3$ . The act's expectation would turn out to be a compound of the form  $PE, PCE = u_1p_1 + u_2p_2, p_3$ . This approach would have the advantages of retaining the distinctions among the different senses of expectation and permitting comparative judgments among the fully comparable, utility-comparable, and plausibility-comparable components of alternative expectations. In addition, the approach could be defended on grounds of coherence with the rationale for the other three decision rules: "Use comparable data!" But how successful such an approach might be is unexplored territory, so far as I know. It must be left here as a topic for further research.

## References

- Allais M (1953) Fondements d'une théorie positive des choix comportant un risque et critique des postulats et axiomes de l'école américaine. *Econometrica* 40:257–332 (trans: Allais, 1979a)
- Allais M (1979a) The foundations of a positive theory of choice involving risk and a criticism of the postulates and axioms of the American school. In: Allais M, Hagen O (eds) *Expected utility hypotheses and the Allais paradox: contemporary discussions of decisions under uncertainty with Allais' rejoinder*. D. Reidel, Dordrecht/Boston/London, pp 27–145 (trans: Allais, 1953)
- Allais M (1979b) The so-called Allais paradox and rational decisions under uncertainty. In: Allais M,

- Hagen O (eds) Expected utility hypotheses and the Allais paradox: contemporary discussions of decisions under uncertainty with Allais' rejoinder. D. Reidel, Dordrecht/Boston/London, pp 437–681
- Amihud Y (1979) Critical examination of the new foundation of utility. In: Allais M, Hagen O (eds) Expected utility hypotheses and the Allais paradox: contemporary discussions of decisions under uncertainty with Allais' rejoinder. D. Reidel, Dordrecht/Boston/London, pp 149–160
- Aumann R (1962) Utility theory without the completeness axiom. *Econometrica* 30:445–462
- Baumann P (2005) Theory choice and the intransitivity of “is a better theory than”. *Philos Sci* 72:231–240
- Black M (1985) Making intelligent choices: how useful is decision theory? *Dialectica* 39:19–34
- Byron M (ed) (2004) Satisficing and maximizing: moral theorists on practical reason. Cambridge University Press, Cambridge
- Chu FC, Halpern JY (2004) Great expectations. Part II: generalized expected utility as a universal decision rule. *Artif Intell* 159:207–229
- Chu FC, Halpern JY (2008) Great expectations. Part I: on the customizability of generalized expected utility. *Theory Decis* 64:1–36
- de Finetti B (1937) La prévision, ses lois logiques, ses sources subjectives. *Annales de l’Institut Henri Poincaré* 7:1–68 (trans: (1980) Foresight: its logical laws, its subjective sources. In: Kyburg HE, Smokler HE (eds) Studies in subjective probability. R. E. Krieger, New York, pp 53–118)
- Ellsberg D (1961) Risk, ambiguity, and the Savage axioms. *Q J Econ* 75:643–669
- Elster J (1979) Ulysses and the sirens: studies in rationality and irrationality. Cambridge University Press, Cambridge
- Fishburn PC (1991) Non-transitive preferences in decision theory. *J Risk Uncertain* 4:113–134
- Friedman N, Halpern JY (1995) Plausibility measures: a user’s guide. In: Besnard P, Hanks S (eds) Proceedings of the eleventh conference on uncertainty in artificial intelligence (UAI ‘95). Morgan Kaufmann, San Mateo, pp 175–184
- Gärdenfors P, Sahlin N-E (1982) Unreliable probabilities, risk taking, and decision making. *Synthese* 53:361–386
- Giere RN (1985) Constructive realism. In: Churchland PM, Hooker CA (eds) Images of science. University of Chicago Press, Chicago/London, pp 75–98
- Good IJ (1962) Subjective probability as the measure of a non-measurable set. In: Suppes P, Nagel E, Tarski A (eds) Logic, methodology, and the philosophy of science. Stanford University Press, Stanford, pp 319–329
- Gowans CW (ed) (1987) Moral dilemmas. Oxford University Press, New York/Oxford
- Greenspan P (1983) Moral dilemmas and guilt. *Philos Stud* 43:117–125
- Halpern JY (2003) Reasoning about uncertainty. MIT Press, Cambridge, MA
- Hammond PJ (1988) Consequentialist foundations for expected utility theory. *Theory Decis* 25:25–78
- Hintikka J (1970) On semantic information. In: Hintikka J, Suppes P (eds) Information and inference. D. Reidel, Dordrecht, pp 3–27
- Hintikka J, Pietarinen J (1966) Semantic information and inductive logic. In: Hintikka J, Suppes P (eds) Aspects of inductive logic. North-Holland, Amsterdam, pp 96–112
- Hirshleifer J (1965) Investment decision under uncertainty – choice theoretic approaches. *Q J Econ* 79:509–536
- Hughes RIG (1980) Rationality and intransitive preferences. *Analysis* 40:132–134
- Jeffrey RC (1983) The logic of decision, 2nd edn. University of Chicago Press, Chicago/London
- Klir GJ (2006) Uncertainty and information: foundations of generalized information theory. Wiley, Hoboken
- Knight FH (1921) Risk, uncertainty and profit. Hart, Schaffner & Marx, Boston. Reprint (1971) University of Chicago Press, Chicago/London
- Kyburg HE (1961) Probability and the logic of rational belief. Wesleyan University Press, Middletown
- Kyburg HE (1979) Tyche and Athena. *Synthese* 40: 415–438
- Kyburg HE (2006) Vexed convexity. In: Olsson EJ (ed) Knowledge and inquiry: essays on the pragmatism of Isaac Levi. Cambridge University Press, Cambridge, pp 97–110
- Levi I (1967) Gambling with truth: an essay on induction and the aims of science. Alfred A Knopf, New York
- Levi I (1974) On indeterminate probabilities. *J Philos* 71:391–418
- Levi I (1986) Hard choices: decision making under unresolved conflict. Cambridge University Press, Cambridge
- Maher P (1993) Betting on theories. Cambridge University Press, Cambridge
- Morgenstern O (1979) Some reflections on utility. In: Allais M, Hagen O (eds) Expected utility hypotheses and the Allais paradox: contemporary discussions of decisions under uncertainty with Allais’ rejoinder. D. Reidel, Dordrecht/Boston/London, pp 175–183
- Ok EA (2002) Utility representation of an incomplete preference relation. *J Econ Theory* 104:429–449
- Ok EA, Dupra J, Maccheroni F (2004) Expected utility theory without the completeness axiom. *J Econ Theory* 115:118–133
- Pollock JL (2006) Thinking about acting: logical foundations for rational decision making. Oxford University Press, Oxford

- Resnik MD (1987) Choices: an introduction to decision theory. University of Minnesota Press, Minneapolis
- Savage LJ (1972) The foundations of statistics, 2nd edn. Dover, New York
- Simon H (1982) Models of bounded rationality. MIT Press, Cambridge, MA
- Slote M (1989) Beyond optimizing: a study of rational choice. Harvard University Press, Cambridge, MA
- von Neumann J, Morgenstern O ([1944]1953) Theory of games and economic behavior, 3rd edn. Princeton University Press, Princeton
- Weirich P (2004) Realistic decision theory: rules for nonideal agents in nonideal circumstances. Oxford University Press, Oxford
- Zimmerman MJ (1996) The concept of moral obligation. Cambridge University Press, Cambridge



# 22 Social Influences on Risk Attitudes: Applications in Economics

*Stefan T. Trautmann<sup>1</sup> · Ferdinand M. Vieider<sup>2</sup>*

<sup>1</sup>Tilburg University, Tilburg, The Netherlands

<sup>2</sup>Ludwig-Maximilians-Universität München, Munich, Germany

<i>Introduction</i> .....	<b>576</b>
<i>History</i> .....	<b>578</b>
<i>Current Research</i> .....	<b>581</b>
The Decision Maker Observes Other Agents' Outcomes .....	581
The Decision Maker's Outcomes Are Observed by Others .....	586
The Decision Maker's Choices Determine or Influence Other Agents' Outcomes .....	588
The Decision Maker's Outcomes Depend on Other Agents' Choices .....	591
<i>Further Research</i> .....	<b>594</b>

**Abstract:** Economic research on risk attitudes has traditionally focused on individual decision-making issues, without any consideration for potential social influences on preferences. This has been changing rapidly over the last years, with economists often taking inspiration from earlier psychological research in their increasing consideration of social aspects in decision-making under risk. We provide a broadly conceived overview of the recent literature, defining four different categories of social influences on economic decisions under risk: (1) the observation of other agents' outcomes; (2) the observation of the decision maker's outcomes by other agents; (3) the direct effect of the decision maker's choices on other agents' outcomes; and (4) the direct dependency of the decision maker's outcomes on other agents' choices. While many promising insights have been gained over the last few years, several shortcomings and inconsistencies in our current understanding of social influences on decision-making under risk are pointed out. The overview concludes with a discussion of two real-world applications – agency in financial markets and climate change – that prominently show the importance of furthering our knowledge in this area. In order to achieve such increased knowledge, a much deeper integration of currently dispersed disciplinary knowledge in the social sciences seems crucial.

## Introduction

---

Decision-making under risk has traditionally been considered an *individual* decision problem in the economics literature. The decision maker considers the outcomes in the uncertain possible future states of the world for different courses of action available, and then chooses the alternative that maximizes some function of the outcomes and the probabilities involved. While individual preferences will always be central for decisions under risk, we argue that *social influences* are also crucial in many economic decisions. Most economic decisions involve situations where agents interact in markets, or strategic situations where the decision maker's own actions affect others, and, vice versa, her actions are affected by others' choices. Even when an agent's outcomes are not directly affected by, or do not directly affect, others, more subtle social issues may play a role. Indeed, the observation of others as well as the awareness of (potentially) being observed by others may influence the agent's actions. Nevertheless, social influences on the evaluation of risky alternatives have not been studied or formalized by economists until recently.

Before delving any deeper into the topic, it appears imperative to introduce some elementary distinctions and definitions. While the concept of risk is taken in a broad sense in keeping with the spirit of this handbook, it may not always be obvious what studies can be considered as *economic* in a sprawling social science literature in which traditional boundaries between disciplines are becoming increasingly blurred. In first approximation, we can define as economic studies all those that have as their primary object of study transactions that affect an agent's wealth: monetary and non-monetary. This, in turn, entails a necessary focus on *outcomes* of decisions – as opposed, for instance, to the focus on the decision-making *processes* themselves, typical of the psychology literature. Indeed, this explains at least in part the different conceptions of rationality that serve as normative benchmarks in different disciplines. While these issues also produce methodological differences, we will be as inclusive as possible, considering any studies on decisions that may result in the creation or transfer of value or wealth.

In this chapter, we will first illustrate the traditional neglect of social issues in risky decisions in economics, and discuss some earlier contributions in neighboring fields, especially

psychology, that have influenced the more recent attempts in economics to model decisions under risk as a social decision task (section [History](#)). We then turn to current research in economics. Various social aspects of risky decisions have received attention in recent years, and we suggest a classification within which to discuss the different approaches and organize the quickly growing literature. We distinguish four broad classes of social influences, based on the relationship between the parties involved and the means by which the influence is transmitted. Each of these classes subsumes different theoretical and empirical approaches. The four classes can be summarized as follows:

1. The decision maker observes other agents' outcomes.
2. The decision maker's outcomes are observed by other agents.
3. The decision maker's choices affect other agents' outcomes.
4. The decision maker's outcomes depend on other agents' choices.

In the first class, where the decision maker observes other agents' outcomes and her own choices depend on this observation, there is generally no direct interdependency between agents. For instance, Paul's choice between a job with a sure annual income of \$50K and a job with a risky income between \$40K and \$65K may be influenced by whether Peter, his brother-in-law, earns \$48K or \$52K. Such social influence is purely psychological, and involves factors like fairness, social reference points, social regret, or conformity. A similar, yet more explicit, situation may be the one in which an agent is given explicit advice on actions to take (e.g., investment advice). Both experimental studies and theoretical models have been advanced to include these factors in models of risky decisions.

The second class, encompassing cases in which the decision maker's outcomes are observed by others, also concerns psychological mechanisms, and mainly relates to *accountability* – the implicit or explicit expectation on the side of the decision maker that she may have to justify her decisions to somebody else. For instance, a risky investment that looks attractive from an individual decision perspective may be less acceptable when anticipating the possibility that negative outcomes are observed by others. To the extent that people care about their reputation, a potential observer's possible judgment of the agent's decision making competence becomes a crucial aspect in the decision between risky alternatives. This, in turn, may lead to deviations from the type of behavior observed in the isolated and purely individual decisions typical of laboratory experiments, in which subjects are assured as much anonymity as possible. Since such isolated decisions will rarely be found outside the economist's or psychologist's laboratory, a thorough understanding of the impact of such social influences seems crucial for our ability to generalize laboratory findings to the real world.

The third and the fourth class of influences refer to situations in which there exists true interdependence between agents, which goes beyond their influences on prices and the functioning of markets. Normatively, these situations may still be, and have been, studied as an individual decision problem. Both empirical research and recent real-world events strongly suggest, however, that explicitly modeling the social aspects of these decisions is warranted. The third class summarizes situations where an agent's decision affects not only her own outcomes, but also those of another agent or group of agents. Financial intermediation and delegation of decision-making power when ownership of assets is diffuse constitute an important class of situations in real-world decision making under risk, for which many problems have been revealed during the recent financial turmoil. Economists have become interested in the question how risk taking on

behalf of others differs from risk taking for oneself, and have begun to study these situations in more controlled environments that allow for the examination of different sorts of influences, which may all be important in the real world.

The fourth class of situations looks at the other side of the relationship in the third class, with the decision maker assuming the role of a principal delegating the final decision power to an agent. This class of influences is also important in strategic interactions and for the empirical analysis of game theoretical models, where agents move simultaneously and hold beliefs about the likelihood of the opponent's strategies. Empirical research has shown that in such situations, choices deviate from those in which the agent faces the same outcomes and probabilistic beliefs, in which however the uncertainty is due to nature rather than another agent. Findings in this category also concern the kind of incentives that a principal can put in place in order to make the agent's decisions coincide as much as possible with her own preferences. This principal–agent issue has attracted much interest in economics, and a host of models, as well as an increasing number of empirical studies, exist on how the risk attitude of agents can be influenced by the principal.

The state of the economics literature on social influences in risk taking is summarized under the Current Research heading according to the four classes of situations just described. While this conceptual separation will be maintained as much as possible, this is done for expositional clarity and it should be well understood that the four classes of influences will not be strictly separate in empirical applications. In particular, in the latter two classes with true interdependence, the purely psychological motives deriving from observation of other agents, or by other agents, will also affect decision making, leading to interesting interactions of motives. When applicable, we will discuss research studying such interaction effects.

In the final section of this chapter, open questions will be identified, as well as applications of the new insights on social influence to economic problems (in the broadest sense) involving risk. Particular attention is devoted to the issue of agency in risky decision making, with a focus on financial decisions. The latter involves all four classes of social influences and has not been widely explored so far. We will point out applications to common resources problems in environmental and climate policy. This section will also discuss how the related fields that provided the initial insights inspiring the study of social influences in economics may contribute additional inputs to improve the first generation of models and empirical studies in economics.

Finally, a warning is in place. While we have attempted to provide a broad overview of the existing economics literature on social influences, this study cannot aspire to being comprehensive. The speed at which new research emerges, as well as the space limits imposed by the current format, forced us to make some difficult choices both on the number of studies to be included and on the level of detail to be reported on those studies. This chapter is thus driven at least in part by the particular interests of the authors, as well as the desire for providing a wide-spun overview rather than an in-depth analysis of some particular aspects.

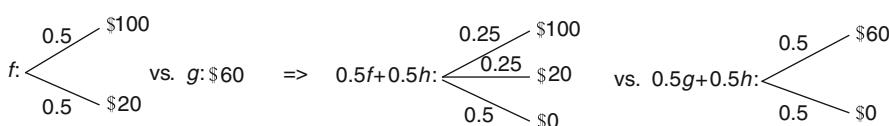
## History

Economists have long studied risk in the tradition of Bernoulli (1738), who introduced the idea of expected utility. A prospect (or lottery) representing a probability distribution over possible outcomes is evaluated by a weighted sum of a function of the outcomes, often called utilities, the weights being the probabilities with which the different outcomes are expected to occur

(expected utility theory) (Chap. 15, A Rational Approach to Risk? Bayesian Decision Theory and Chap. 5, The Economics of Risk: A (Partial) Survey). In the mid-twentieth century, behavioral foundations were laid for the expected utility model that defined the model in terms of axioms describing simple observable preferences between risky prospects (von Neumann and Morgenstern 1947; Herstein and Milnor 1953). For instance, the so-called independence condition has proved crucial for preferences to be describable by expected utility theory. It stipulates that the preference between two prospects  $f$  and  $g$  should not be affected if we probabilistically mix both of these prospects with a third prospect  $h$ . (Figure 22.1) illustrates the independence condition.

Foundations have also been given for the technically distinct case in which no objective probabilities are known, a situation often referred to as *uncertainty* (Knight 1921). In this case, the decision maker evaluates a prospect by using her own, subjective estimates of the relevant probabilities (Anscombe and Aumann 1963; Savage 1954). Such behavioral foundations of decision-making models have typically focused on observable individual preferences, and have made no reference to social influences. More generally, dependence of decisions on a particular context has usually been excluded from the economic definitions as well as being avoided in empirical investigations, because it may give rise to ad hoc explanations and lack of predictability. In the case of social context, a model that allows for different behavior for each possible social context was long considered unfalsifiable and empirically useless.

Remarkably, social influences have also been excluded when studying situations explicitly concerned with social decisions. The well-known foundation of utilitarianism laid by Harsanyi (1955) employs individual agents' expected utilities over risky distributions of their own social positions to derive the social planner's utilitarian welfare function. While explicitly concerned with social comparisons, individual evaluations of risky social allocations are assumed to exclude any social comparisons (Diamond 1967; Trautmann 2010). Similarly, game theoretic analyses of strategic interactions involve both risk due to nature and risk due to the decisions made by other players. No distinction between such different types of risks has been made until recently, even though these two kinds of risks may be perceived very differently by the decision maker (Fox and Weber 2002). Von Neumann and Morgenstern (1947) introduced their expected utility model of individual risk taking to analyze strategic game situations and mixed equilibria, where randomization of strategies by other players introduces an element of risk into each player's decision problem. As we will show in the next section, the economics literature has only recently started to capture the systematic aspects of social influences and to improve the realism of such models, while maintaining predictive power and empirical refutability.



**Fig. 22.1**

**Mixing and independence.** Notes: Independence stipulates that the preference between the left-hand prospects  $f$  and  $g$  should not be affected if both prospects are probabilistically mixed with the same prospect  $h$ . The figure shows a 0.5 mix with a prospect  $h$  resulting in a sure zero payoff

In contrast to the economics literature, psychologists have long been aware of social influences and tried to take these explicitly into account in their theorizing. Indeed, the issue of social influences is important enough as to have given rise to the whole subfield of social psychology. A host of different forms and shades of social influences have thus systematically been explored in social psychology, ranging from the conformity paradigm (Asch 1955), to social facilitation or mere presence (Bond and Titus 1983; Zajonc 1965), accountability in its various forms (Lerner and Tetlock 1999; Shafir et al. 1993), and many more (Cialdini 1993). This long research tradition on social influences has produced over the years a huge pool of knowledge from which economists can draw inspiration. One aspect of this research tradition that has complicated the interdisciplinary exchange is the divisive nature of many of the research efforts in social psychology, often aimed at further subdividing established effects rather than at searching for commonalities between existing research paradigms. Such aggregation could indeed benefit the dialogue between disciplines in the social sciences. Furthermore, while a huge body of evidence has been produced on social influences on decisions in general, the case of decisions under risk appears to play a relatively minor role in this tradition.

There are exceptions, however, and a non-negligible body of psychological evidence on social aspects influencing behavior under risk has emerged over the years. A prominent example is *risky shift*, which is defined as a level of risk tolerance in a group that is larger than the average level of individual risk tolerance of its members (Stoner 1961). Wallach et al. (1964) examine the causes of the risky shift phenomenon, and find that the diffusion of responsibility in the group leads to increased risk taking. They also find the opposite effect, a *cautious shift*, however. The latter occurs if individual members are personally responsible for the welfare of the group. A very similar result is obtained by Weigold and Schlenker (1991), who examine these effects through the lens of the social facilitation paradigm. Having elicited subjects' risk attitude, they found that subjects became more extreme in their original positions when they were told that they would be required to justify their choices, with risk-averse subjects becoming more risk averse, and risk seekers increasing their preference for risk. Below, we show that there is converging evidence from the more recent economics literature that responsibility for others often leads to reduced risk tolerance, as would be predicted if people are predominantly risk averse in isolated decisions.

Other approaches in psychology have been closer to economics in the sense that their focus centered on individual preferences, accounting for individual-level biases. Kahneman and Tversky's (1979) prospect theory, for instance, unmasks various deviations in risky choices from those predicted by expected utility theory (● Chap. 25, Risk Perception and Societal Response and ● Chap. 19, Paradoxes of Rational Choice Theory). These deviations, however, are explained through individual-level psychophysical aspects like diminishing sensitivity, reference dependence, and a larger impact of negative outcomes (losses) than of positive outcomes (gains), in comparison to some reference points. Nevertheless, social aspects can be incorporated naturally into this framework, for instance, through the undetermined origin of the reference point or various editing processes. Boles and Messick (1995) showed that the selection of the decision maker's reference point in the presence of multiple possible reference points depends on social comparisons. These authors find that social reference points are often more influential than individual reference points, like the status quo. In one experiment, compared to the status quo of \$0, people preferred a gain of \$90 when another person simultaneously received a gain of \$500 less than a much smaller gain of \$10 when the other person experienced a loss of \$100.

Psychologists have also studied how the anticipation of being evaluated by others can affect attitudes under uncertainty. We have already mentioned Weigold and Schlenker's (1991) result showing how accountability amplifies pre-existing risk attitudes, making risk averters more risk averse and risk seekers even more risk prone. Mere formulation differences or frames have often been found to influence decisions under risk and uncertainty. Takemura (1993, 1994) showed that making subjects accountable by announcing that they will have to justify their decisions in front of somebody else reduced incoherence between different frames, thus increasing the consistency of preferences. Similar effects were obtained by Miller and Fagley (1991) and Sieck and Yates (1997). Studying a problem in which subjects' decisions had consequences for others, Tetlock and Boettger (1994) found that subjects were more reluctant to approve a risky drug for a hypothetical market when held accountable. Indeed, accountability pushed subjects to procrastinate as well as to try to pass the responsibility on to others, so that potentially useful drugs were not approved for use in a timely manner because of the idiosyncratic risks they entailed.

Ambiguity aversion – the preference of known-probability outcome generating processes over normatively equivalent processes entailing unknown probabilities (Fox and Tversky 1995; Frisch and Baron 1988) – is a classic example of how economists often arrive late to the study of social-decision aspects. The phenomenon itself has been of interest to economists ever since the publication of the famous Ellsberg paradox in 1961 (☞ Chap. 18, [Unreliable Probabilities, Paradoxes, and Epistemic Risks](#)). While economists have studied the issue as a purely individual problem, psychologists have taken a wider approach to the issue. Curley et al. (1986) studied a wide array of potential causes of ambiguity aversion. They found that being observed by a group of people when making a choice between a bet on a known-probability prospect and a bet on an ambiguous prospect, decision makers became more ambiguity averse (a similar finding was reported by Taylor (1995)). Economists became interested in social influences on ambiguity attitudes only very recently. This parallels developments in the literature on the effects of social reference points and other influences, in which economists did not develop an interest until more than a decade later than psychologists. For this reason, the psychological findings often form the basis for economic studies. It should be noted however that – in addition to having a different focus deriving from the different underlying research paradigm – economists studying social influences in decisions under uncertainty are not always aware of previous work by psychologists. In the following section, we will therefore also refer to findings in the psychological literature that are of direct relevance to problems studied more recently in economics.

## Current Research

---

In this section, we review current research in economics. We follow the structure laid out in the introduction and discuss any overlaps between the classes of influence where applicable.

### The Decision Maker Observes Other Agents' Outcomes

---

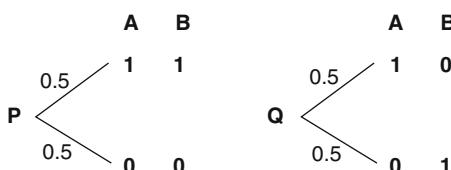
A variety of social influences fall into this class. We discuss three different types of models. First, fairness motives have been shown to interact with uncertainty if the decision maker observes the payoffs of other agents whose outcomes she considers relevant. Second, the

outcome of another person may act as a social reference point even in the absence of fairness considerations. And third, conformity with the behavior of their peers has been shown to affect people's decisions under risk.

**Process Fairness.** Fairness motives may alter a decision maker's risk attitude when other people's decisions and outcomes are observable. Consider the risky decision for person A shown in [Fig. 22.2](#). Person A faces a choice between two risky lotteries P and Q. For person A, each lottery gives a payoff of 1 or a payoff of 0, with equal probability of 0.5. If player A considered only her own payoffs in her decision, she would be indifferent between the two lotteries – indeed, they would be exactly the same. If, however, the outcome of person B is observed and equality concerns matter to person A, then lottery P will be strictly preferred to lottery Q. For either state of the world, equality obtains for prospect P and inequality obtains for prospect Q, even though the expected value of the two prospects is equal for both agents.

The importance of such fairness motives in risky decisions was illustrated empirically by Kroll and Davidovitz (2003), Bolton et al. (2005), and Krawczyk and Le Lec (2010). Zizzo (2004) and Cappelen et al. (2009), on the other hand, show the inverse effect, namely that individual risk-taking influences fairness evaluations. Interestingly, the interaction of risk and fairness has also led to the new concept of *process fairness* in economics. Process fairness explicitly considers the risk borne by all agents and does not evaluate the equality of outcomes, but the equality of expected outcomes. In the above example, person A is indifferent between prospects P and Q if she is motivated by process fairness because her own payoff is identical under both prospects, and the expected payoffs of both agents are equal and identical under both prospects. Theoretical models accounting for allocation of risk and process fairness have been proposed by Bolton et al. (2005), Trautmann (2009), Sebald (2010), Krawczyk (2011), and Borah (2010). Trautmann (2010) discusses an application of individual-level process fairness to social choice problems involving risk. Bolton and Ockenfels (2009) report experimental evidence for the relevance of process fairness considerations in choice situations similar to those in [Fig. 22.2](#), where person A makes choices for herself and for a passive person B (see below). Aldashev et al. (2010) show process fairness effects in real-world settings. In the risky allocation of more or less desirable tasks to workers, they show that process fairness influences the effort that these workers exert in their task.

Social influences due to process fairness concerns can be employed to justify the concept of *resolute choice* in risky decisions, introduced by Machina (1989) to restore dynamic consistency of non-expected utility preferences. If their choices violate the independence axiom, non-expected utility maximizers may change their preference after the resolution of the uncertainty, violating dynamic consistency. Trautmann and Wakker (2010) show that process fairness preferences imply a similar rejection of consequentialism and lead to dynamic inconsistency. In [Fig. 22.2](#), process fairness implies indifference between prospects P and Q. Let us assume



**Fig. 22.2**

Risky decision for Person A who observes the risky outcome of Person B

that Q is chosen and resolved. Obtaining the poor outcome under prospect Q after the resolution of uncertainty, however, the agent faces a sure allocation and may reconsider her preference. In particular, she may now prefer the poor outcome under prospect P rather than prospect Q, because of ex-post fairness concerns. Under process fairness, however, resoluteness becomes convincing, because agents explicitly consider risks borne in the past and will not adjust their preferences ex-post. Trautmann and Wakker (2010) argue that process fairness can thus give a justification of resolute risky choice in social settings. Under process fairness, the risks borne in the past by the players become meaningful in terms of social comparison and should be considered also ex-post, enforcing dynamic consistency.

*Social reference points.* A related but slightly different view holds that another agent's outcome may act as a reference point in a risky choice, similarly as in the psychological work by Boles and Messick (1995). Linde and Sonnemans (2009) and Rohde and Rohde (2009) provide theoretical models and empirical tests of this paradigm. Consider prospect P in [Fig. 22.2](#), but now assume that agent B receives an outcome of at least 1 in both states (instead of the given numbers). Linde and Sonnemans call this a social loss frame for agent A. On the other hand, if B receives an outcome smaller than 0 in both states, the prospect is evaluated in a social gain frame for agent A. The situation depicted in [Fig. 22.2](#) for prospect P, with a stochastic outcome for B equal to the outcome of A is called a neutral frame. In contrast to fairness-based influences on risk attitude, here the social comparison is thought to influence risk attitude by providing a reference point. Building on prospect theory's reference-dependent evaluation function, risk seeking is predicted in loss frames while risk aversion is expected to prevail in gain frames.

The empirical evidence does not support this prediction unequivocally, however. Linde and Sonnemans observe more risk aversion when decision makers find themselves in a social loss frame than under the gain frame, reversing the typical prospect theory pattern. In contrast, in a field experiment devised to test this idea, Haisley et al. (2008) found that poor people are more likely to buy lottery tickets after being shown income brackets that place them at the bottom of the income distribution, as compared to when the income brackets employed place them somewhere in the middle ground. That is, these people become more risk seeking when their placement in a perceived peer group puts them in a social loss frame. These different findings may be explained by the role of the reference point as an aspiration level (Diecidue and van de Ven 2008). People put additional value on outcomes that help them to achieve their aspiration level, for instance, a social comparison outcome. While a simple loss frame may not elicit risk seeking per se, it may do so if the aspiration level can be met or surpassed by assuming greater risks, which is surely the case for the lottery ticket in the Haisley et al. study.

A related concept, introduced by Delgado et al. (2008), is social loss aversion. Delgado et al. studied social loss aversion for first-price auctions in the presence of multiple social reference points. Consider once more prospect Q in [Fig. 22.2](#). This allocation lottery can be interpreted as a situation where either person A or person B receives a desirable item (with 1 indicating the person obtaining the item). Auctions provide an example of such situations, with uncertainty deriving from other bidders' behavior. The auction winner receives the item while all other bidders receive nothing. Bidding behavior in auctions has widely been studied in economics, both theoretically and empirically. An empirical phenomenon that has attracted much interest is the fact that people tend to overbid in first-price auctions with respect to the theoretical prediction of the Nash equilibrium. In the first-price sealed bid auction, each bidder simultaneously submits a bid, and the winner with the highest bid wins the auction and has to pay her

bidding price. To make a positive profit, the buyers have to submit bids that are lower than their valuation of the good. The lower the bid submitted, the higher the profit but also the higher the risk that another bidder will submit a larger bid. Nash equilibrium predicts the optimal tradeoff between these two forces, but empirically people usually bid too high relative to the Nash prediction. Risk aversion has sometimes been used to explain this overbidding. More recently, using both behavioral and neuroeconomic evidence, Delgado et al. (2008) showed that it is mainly the perception of non-winning of the auction (lower branch of prospect Q) as a *social loss* that influences overbidding.

These examples show that straightforward extrapolation from individual to social reference points is not always warranted. Some individual effects may be amplified in social contexts as was the case with loss aversion; others may be reversed as in the risk aversion for social loss frames. Furthermore, by moving from individual to social reference points, complexity increases because of the multiplying of reference points (individual and social reference points usually coexist), and stochastic reference points if other agents also face risky prospects resulting in different outcomes in different states of the world. The development of descriptive models with social reference points will therefore have to draw upon psychological insights about the aggregation of multiple references points (e.g., Kahneman and Miller 1986) and theoretical work in economics on stochastic reference points (Sugden 2003; Schmidt et al. 2008).

*Conformity and Peer Effects.* Recent research has shown that conformity with the behavior of relevant others is an important driver of economic decisions (Sacerdote 2001; Duflo and Saez 2002; Falk and Ichino 2006). An early theoretical model is provided by Bernheim (1994). There is no clear-cut evidence in economics on the effect of conformity on risky decisions yet. Goeree and Yariv (2006) study situations where agents have to predict the distribution of colors in an urn after either observing one sample from the urn or observing other agents' predictions. In particular, either an urn with three red and seven blue balls, or an urn with seven red and three blue balls is randomly selected, but subjects do not know which urn was chosen. Before predicting the distribution, subjects can either draw a sample ball with replacement from the selected urn at zero cost, or observe the betting choices of four other persons who had no information before making their bet. That is, while a sample draw is informative from a Bayesian viewpoint, observing the uninformed choices of the other four people does not reveal any information about the distribution in the urn. Goeree and Yariv observe clear preferences for observing the uninformed choices of others. Between 30% and 50% of the subjects preferred observing others' choices, and this pattern is robust if the authors control equality and efficiency concerns. Further, conformity with the uninformed majority choice becomes stronger for larger majorities, consistent with the psychology of social impact (Latane 1981). Goeree and Yariv therefore interpret their result as a preference for conformity.

While this interpretation seems the obvious explanation in the simple setting used by Goeree and Yariv, the result may obtain even under rather weak conformity preference if subjects have no clear concept of the Bayesian learning from sampling with replacement. A recent study by Trautmann and Zeckhauser (2010) indeed suggests that many people have no proper understanding of the concept of learning under uncertainty and do, consequently, often completely neglect learning opportunities. If people are offered a chance to observe others, they may choose to do so even if conformity preferences are weak, or out of pure curiosity.

Another study finding strong peer effects was conducted by Cooper and Rege (2008). These authors asked subjects to make a set of simple choices between risky prospects. The whole set of choices was made three times. In the first instance, subjects made simple individual choices.

In the second and third repetition, they were either informed only about their own past choices or additionally about the past choices of five other people in their group. The authors found strong effects of the additional peer group information on risk attitudes. Two theoretical accounts were put forth: one based purely on conformity pressures, and another one based on social regret. Under pure conformity, the option chosen initially by a majority of peers yields an additional social utility component, such that groups are predicted to converge toward the initial majority choice over time.

The social regret model for peer effects on the other hand is based on the assumption that decision makers anticipate regret in risky decisions (Loomes and Sugden 1982; Bell 1982). Consider the right-hand situation Q in  Fig. 22.2, and assume that A and B represent the outcomes of two distinct prospects. If a person has to decide between prospects A and B, she may anticipate that after choosing prospect A she feels regret in case the lower branch should obtain. In this case, prospect B would have been the better choice ex-post. Similarly, after choosing prospect B, the upper branch outcome would lead to decision regret.

In many situations, prospects are asymmetrically affected by regret, which then strongly influences choice behavior. Cooper and Rege assume that subjects evaluate risky prospects by an evaluation function that consists of an “inherent utility” term, for instance, the expected utility of the prospect, and an additive punishment for anticipated regret. Importantly, the weight of the regret term in the evaluation inversely depends on the expected number of people choosing the same gamble – misery loves company. This model is similar in spirit to the above-discussed amplification of individual-level biases in social circumstances.

Studying the dynamic behavior in their experiment, Cooper and Rege find evidence for the social regret model of peer effects and reject the pure preference for conformity model. There is no simple convergence toward the initial majority as predicted by conformity, but rather a bias against more risky prospects in the peer choice information treatments compared to the individual information-only treatments. This bias is predicted by social regret, because the risky options involve more potential for regret than safer options. Thus, in contrast to Goeree and Yariv (2006), this study does not find a pure conformity effect in risky choices. Another study that does not support conformity is that of Corazzini and Greiner (2007). In their study, subjects choose between two identical gambles A and B one after the other. That is, by definition, individual attitudes toward risk cannot imply a strict preference between the gambles, allowing conformity to potentially exert as strong an influence as possible. Subjects making decisions later in the sequence should thus tend to the majority choice of subjects who previously revealed their preference. Corazzini and Greiner report significant non-conformity: observing a sequence of choices of one type by other subjects, subjects are more likely to choose the other prospect. This pattern is also observed if prospect A is strictly better than prospect B. Still about 25% of the subjects prefer to choose B after observing a sequence of A choices. Corazzini and Greiner’s result is consistent with earlier evidence in psychology reported by Arkes et al. (1986), who executed a experiment on learning under uncertainty in which subjects were given a decision-making rule, which would allow them to maximize their payoff. They found that subjects displayed a strong tendency not to follow that rule, but rather tried to beat the system – with the consequence of a substantial loss in payoffs as compared to subjects who did follow the rule.

On the whole, the existing evidence on the role of conformity is thus inconclusive and more research is needed to uncover the reasons underlying the widely differing results. As Cooper and Rege show, it is often difficult to separate different explanations with purely behavioral data. New methods in economics, including data from brain-scans executed while decisions are

taken may help to study such difficult-to-identify social influences. For example, Engelmann et al. (2009) conducted a neuro-economic study of advice in financial decision making under risk. They found that brain regions involved in the value tradeoffs typical for decision making under risk become less activated if advice by an expert is given. On the other hand, regions involved in strategic thinking and mentalizing others' intentions become more active (see also Bossaerts 2009, Sect. 6). This shift in processes results in behavioral changes. Identifying underlying neurological processes can thus help identify changes in social decision-making under risk where behavioral observations are inconclusive.

## The Decision Maker's Outcomes Are Observed by Others

---

Situations in which a decision maker is observed by others, and the effects that such observation may conjure about, has been widely studied in social psychology. Once again, the interest for the topic in economics surfaced much more recently. In early studies of the issues of observability of donations, Hoffman et al. (1996) showed that in a dictator game subjects offered lower sums of money when nobody could observe the amount donated than when donations were to some degree observable. Even under complete anonymity, however, donations remained superior to the predicted zero result implied by the selfishness assumption typical of economics. While subsequently other experiments used similar concepts of social observability (referred to with a variety of names such as justification need, anonymity, and double blind procedures), they were mostly concerned with behavior in interactive games rather than individual decisions under risk (e.g., Bohnet and Frey 1999; Dufwenberg and Muren 2006).

The overall evidence on the effect of being observed on the decision maker's choices under risk is mixed. In an early psychological study, Weigold and Schlenker (1991) found that subjects' original risk attitudes were enhanced when they anticipated the need to justify them – with risk averters becoming more risk averse and risk lovers becoming more risk seeking. The economic consequences of this are difficult to pin down, however. Vieider (2009) found an effect of accountability on loss aversion, which was decreasing when subjects were held accountable. In this case, economic consequences are clear, since loss aversion is commonly perceived as a decision bias that may lead people to pass over potentially lucrative opportunities. Excessive loss aversion is often thought to occur because of narrow bracketing of decision problems, whereby decisions are considered one by one rather than in the larger perspective of lifetime decisions (Read et al. 1999; Rabin and Weizsäcker 2009). Accountability seems effective inasmuch as subjects realize the irrationality of their instinct to avoid losses, and try to correct it when they anticipate having to justify their decision. In contrast, in an experiment on framing, Vieider (2011) presents evidence that accountability may not be effective at changing risk attitudes either in the gain or the loss domain. This is an issue that is conceptually separate from the loss aversion results just discussed, which require mixed prospects over gains and losses to be addressed. These results do, however, contradict the earlier results by Weigold and Schlenker if one accepts that a majority of subjects is typically risk averse for gains. Weigold and Schlenker's results would then lead us to predict an overall increase in risk aversion for gains under accountability. Overall, the verdict is thus not clear.

Accountability, on the other hand, has been found to increase the coherence between different decision frames – thus increasing the alignment of risk attitudes for gains and for losses. In the same spirit, accountability has been found to push decisions closer to the normative economic

prediction when subjects are called to choose between different prospects, one of which is clearly superior to the other. The commonly observed failure of rational decision making in such choice problems derives generally from the differential complexity of the prospects, so that subjects seem to forego potential gains in expected utility to avoid the cognitive effort required for the discovery of the optimal decision (Huck and Weizsäcker 1999; for a psychological perspective see Inbar et al. 2010). Accountability has been found to increase cognitive effort in such situations, and thus to improve decision making (Kruglanski and Freund 1983). Vieider (2011) showed that improvements due to accountability also occur when at the same time financial incentives are provided for the decisions, so that the effect is not due to particularly low cognitive effort when decisions do not carry real consequences. Rather, he found that the provision of monetary incentives may, paradoxically, sometimes impair the decision-making process, pushing them farther away from the benchmark of economic rationality.

As mentioned in the  History section, accountability has also been found to have effects on ambiguity aversion (Ellsberg 1961). Curley et al. (1986) compared a group for which the decision was performed in relative privacy in front of the experimenter only to one in which the decision was performed in front of the experimenter and the whole group of subjects partaking in the experiment. They found a significant increase in ambiguity aversion when a group of subjects could observe the choice and the outcome of the choice. More recently, Muthukrishnan et al. (2009) successfully used the accountability effect to induce increased ambiguity aversion as a treatment variation in a marketing study.

To test whether social influences may in some settings be necessary for ambiguity aversion, Trautmann et al. (2008) implemented the opposite manipulation by completely excluding the possibility of observation of the decision maker's preferences over outcomes, and thus ultimately of whether the decision maker ends up winning or losing. This was done by using outcomes over which the typical decision maker exhibits a clear preference (in this case, two movies on DVD), whilst such preferences cannot be guessed by the experimenter. The authors found that subjects who had to declare their preference before making their decision exhibited the usual pattern of ambiguity aversion, while subjects who did not reveal their preference beforehand were no longer ambiguity averse. On top of the results obtained by Curley et al., this shows on the one hand that the potential observability of outcomes is sometimes necessary for ambiguity aversion to be strong. On the other hand, it also shows a more subtle fact. Indeed, in Curley et al.'s experiments, the size of the audience was increased, while both the decision itself (i.e., the decision-making process) and the outcomes that obtained from that decision were observable in both conditions. Trautmann et al., on the other hand, vary the observability of *outcomes* alone (or of preferences over outcomes to be precise) across conditions, while the decision-making *process* was always observed by the experimenter. This goes to show that the effect is indeed produced by a focus on outcomes.

Field evidence also points in the direction of increased ambiguity avoidance under observability. In financial investment, the advent of online brokerage has provided an environment of increased anonymity compared with traditional forms of brokerage. Barber and Odean (2001, 2002) showed that people invest more heavily in growth stocks and high-tech companies, investments that are associated with higher ambiguity in the finance literature, when they use online brokerage rather than traditional brokerage. Similarly, Konana and Balasubramanian (2005) found that many investors use both traditional and online brokerage accounts, and hold more speculative online portfolios. Consistent with the proposed accountability effect on ambiguity attitude, one of the investors they interviewed noted in the context

of online trading (p. 518): "I don't have to explain why I want to buy the stock." Barber et al. (2003) showed that group decisions (investment clubs) similarly lead to a stronger preference for easy-to-justify investments.

Economists have just begun to study accountability effects in decisions under risk. While so far we have considered purely psychological accountability effects, in economics, accountability often takes the form of incentive structures in agency problems. We discuss risk taking in such situations in more detail in the following sections.

## The Decision Maker's Choices Determine or Influence Other Agents' Outcomes

---

A relatively common situation is the one in which a decision maker's choices and actions determine not only her own outcomes, but also the outcomes of others – beyond any influence they may have on market prices (a fact that has always been studied in economics, and that we do not consider a social issue). This happens for instance whenever the decision maker is hired to make decisions for someone else, or makes decisions for a group of people as in a business decision affecting various stakeholders. This category also includes decision making within the family regarding such issues as saving and investment, pension and life insurances, or buying a house. The social aspects of such decisions are often overlooked. Decisions for others furthermore occur in strategic game settings where the choices of players affect each other's outcomes, and therefore constitute a risk in the sense that choices are made simultaneously and chosen strategies are unobservable *ex-ante*. We first discuss the literature on decisions with responsibility for others' outcomes under risk due to nature, and then under the more implicit risks stemming from strategic interactions in games. Potential differences in behavior between situations of risk due to nature and of risk due to other people's choices in strategic settings will be dealt with in the following section. We do not review group decision making in detail in this section, but focus on situations of responsibility. Group decisions are complex and involve voting rules and dynamic decision processes that make it potentially difficult to identify effects on risk attitudes. For details on group decisions, we refer the reader to Conradt and List (2010) for a broad overview, and to Isenberg (1986) for a focus on risky decisions.

*Risky decisions on behalf of others.* Risky decisions on behalf of others have been studied in a number of different contexts recently, including household and group decisions. An example is the study by Bolton and Ockenfels (2009) on fairness effects on risk taking discussed above. A problem with this type of studies is that the pure effect of responsibility for others often cannot easily be separated from other influences. If people decide for others as well as themselves, fairness or spitefulness issues become important. If people make decisions in groups, the preference aggregation and joint decision process influence behavior irrespective of potential social influences (De Palma et al. 2010; Wallach et al. 1964). We can nevertheless draw some initial conclusions about the effects of responsibility for others from the literature.

Bolton and Ockenfels (2009) included one set of questions in their study where fairness aspects are held constant although the risk affects both the decision maker and another person. They find increased risk aversion in this situation of responsibility relative to the individual benchmark, even though their result falls short of statistical significance. This finding is consistent with evidence from two studies from the financial management literature, which explicitly ask subjects to make decisions for others that do not affect their own outcomes.

Eriksen and Kvaloy (2010) study amounts invested in a risky asset using a task popularized by Gneezy and Potters (1997), and Reynolds et al. (2009) study simple choices between risky and safe lotteries. Both studies find increased risk aversion in decisions for clients compared with the benchmark of decisions for oneself. Interestingly, Eriksen and Kvaloy find myopic loss aversion, a violation of the expected utility model that has been shown for individual investors (Gneezy and Potters 1997), also in decisions for others. This evidence is consistent with findings in Bateman and Munro (2005) that joint household decision making suffers from the same violations of expected utility as individual decisions do, as well as with the finding that professional traders incur this bias more than the typical student populations employed in experiments (Haigh and List 2005).

While these results support the idea of more cautious decision making under conditions of responsibility, there are also studies that obtain the opposite finding of less risk aversion in decisions for others than for oneself. Chakravarty et al. (2010) find more risk seeking by people who make decision for others in lottery choices and bidding in auctions when decisions do not affect the agent's own payoff. Sutter (2009) finds that teams invest more in the Gneezy and Potter's investment task than individuals, thus pointing in an opposite direction than the findings by Eriksen and Kvaloy (2010), although these differences may be affected by the group decision features of Sutter's experiment. Relatedly, Lefebvre and Vieider (2010) found that compensation through stock options made experimental executives take risks that were excessive from their shareholders' point of view, with a substantial loss of revenue for the latter. When executives, however, knew that they might be called to justify their decisions in front of a shareholder reunion, they sacrificed their own payoff to act in the interest of shareholders, whose returns were increased considerably. The latter study differs from the ones described previously on the point that the compensation of agents and principals is determined through different mechanisms – an important class of relationships in the real world, on which little clear evidence exists to date (see [Further Research](#) section for a discussion).

An interesting variant of risk taking for others was studied in Haisley and Weber (2010). These authors offered subjects choices between more or less selfish allocations between themselves and a passive other person, with the latter affected by uncertainty. In particular, the passive person's payoff depended on either a risky or an ambiguous lottery ([Table 22.1](#)).

If subjects choose the other-regarding option, they receive \$2 while the other person receives \$1.50. This decision problem effectively puts a price of \$.25 on the morally more acceptable choice of providing a higher payoff to the other person. It is thus interesting to see whether the type of uncertainty affecting the other person's payoff influences the moral decision. Haisley and Weber reported an increase in selfishness if the probability with which the payoff of the other person obtains is ambiguous rather than a known 50% chance to win \$1, or else nothing.

**Table 22.1**

Social allocations with risk or ambiguity (Adapted from Haisley and Weber 2010)

	Self	Other
Option A (selfish)	\$2.25	Risky lottery (50–50% chance to win either \$0 or \$1.00)
		Ambiguous lottery (unknown probabilities of winning either \$0 or \$1.00)
Option B (other- regarding)	\$2	\$1.50

They found evidence that in the ambiguous situation, the decision maker assumes an optimistic interpretation of the risk, thus in fact making a tradeoff between (\$2, \$1.50) and (\$2.25, “most likely” \$1.00). Under the known 50% chance to receive nothing, the moral constraint is much stronger because self-serving interpretations are not as easy in that case. Interestingly, if subjects first make a simple choice between an ambiguous and a risky gamble of the kind proposed by Ellsberg (see [History](#) section) before making the social allocation choice, the effect is eliminated. In the initial choice, many subjects reveal ambiguity aversion, thus lower evaluations of an ambiguous chance to win \$1.00. The self-serving positive interpretations thus cannot as easily be constructed in the subsequent moral decision.

*Strategic decisions for others.* A different class of phenomena is the one of decisions that affect others, whereby the risk derives from strategic uncertainty in games (this form of social risk is discussed in more detail also below for the case of individual decision makers). Gong et al. (2010) and Sutter et al. (2010) studied how individuals versus teams act in games in which they had to choose between a low payment Nash equilibrium strategy and a more profitable, but also more risky, non-equilibrium strategy. Both papers find more coordination, and thus risk seeking, for individuals than for groups. As discussed before, from group decisions we cannot easily draw conclusions regarding social influences on risk preferences per se, because of the influence of the joint decision-making process. Charness and Jackson (2009) study coordination in Rousseau’s famous stag hunt game by players who individually choose their strategy, which will affect not only themselves but also the individual paired with them. Because the stag hunt game nicely illustrates how risk obtains from strategic uncertainty, we show the normal form of the game in the following table ([Table 22.2](#)).

As can be seen from the table, both players would prefer to coordinate on hunting stag together. However, decisions are made simultaneously and anonymously. Hence, for each player, it is very costly should the other player choose hunting hare instead, possibly because of spitefulness, or simply because she makes a mistake. Depending on player 1’s beliefs about the probability of player 2 choosing stag, denoted  $p$ , she has to make a choice between a sure payoff of 8 for strategy *Hare* and an uncertain payoff of  $p \times 9 + (1 - p) \times 1$  for strategy *Stag*, thus constituting a decision under uncertainty.

Charness and Jackson let people play the game individually or accompanied by an anonymous and passive dependent other person whose payoffs are identical to the decision maker’s payoffs. They find that people on average become more risk averse, and thus cooperate less, when they are responsible not only for themselves but also for the other person’s payoff. Using within-person observations, they also find that this shift is mainly due to one third of the population who are strongly affected by the responsibility, while two third are not affected at all, suggesting systematic heterogeneity in risky decision making for others. Notice also how

**Table 22.2**  
Payoffs in the stag hunt game (Charness and Jackson 2009)

Player 1 \ Player 2	Stag	Hare
Stag	(9,9)	(1,8)
Hare	(8,1)	(8,8)

( $x,y$ ) refers to player 1’s payoff of  $x$  and player 2’s payoff of  $y$

this effect is consistent with the findings on ambiguity aversion discussed in the previous section, where accountable subjects have been found to shy away from the ambiguous option, which is generally perceived as “riskier” by subjects. The following subsection examines in more detail the effect of risks due to strategic uncertainty if compared with risks caused by neutral events.

## The Decision Maker’s Outcomes Depend on Other Agents’ Choices

---

This class of social influences is concerned with the concept of strategic uncertainty (Van Huyck et al. 1990), in which players’ outcomes depend on the choices of other players. The stag hunt game discussed above is an example. More generally, coordination games where the optimal social outcome provides no incentive for deviation, but players cannot make binding agreements, are a widely studied group of games involving such social risk. In economics, an important application of such coordination games concern depositors’ decisions on whether to keep their money in a distressed bank. If all depositors stay calm, the bank will survive. If a few depositors withdraw their money prematurely, however, the bank will fall and the remaining depositors lose their money. From a game theoretic perspective, all depositors staying calm as well as all depositors withdrawing (a bank-run), both constitute Nash equilibria. Game theory makes no prediction which equilibrium will obtain, and equilibrium selection in such settings therefore creates strategic uncertainty.

Empirical evidence has shown that, although seemingly obvious, the coordination problem is often too difficult for players to solve. After a few rounds of repeated play (with new pairings), the game usually converges to the Pareto-dominated equilibrium of all players withdrawing. The result suggests that strategic uncertainty may be difficult for people to deal with. In a comprehensive study, Heinemann et al. (2009) investigated risk taking in coordination games. Two important results were obtained. First, they found a strong correlation between risk taking in a simple gambling task, and coordination (and thus the acceptance of strategic uncertainty). Second, the authors showed that coordination was more likely the smaller the percentage of people necessary for successful implementation of the Pareto-efficient outcome. Note that because of the multiplicity of equilibria, it is not theoretically the case that a smaller percentage implies a higher chance of an efficient outcome. People seem to perceive it as such, however. Another interesting finding shows that risk-averse people also expect others to be risk averse (see also Kocher and Trautmann (2010) for a similar effect for ambiguity attitude in a market setting). This finding helps explain the poor outcomes of coordination games. In a risky decision, attitudes toward risk should be independent of the perception of the lottery itself. Due to the above-stated correlation, however, in a coordination game, the expected probabilities of the favorable outcome are reduced inasmuch as the decision maker’s own risk aversion creates an expectation of others also shying away from the risk of coordinating on a superior equilibrium.

The conclusion of increased risk aversion under strategic uncertainty compared with the risk due to nature is not unchallenged. Fox and Weber (2002) let subjects choose between either playing a two-player coordination game, or making a risky coin flip. The game is similar to the one described above with two pure equilibria, and with a mixed equilibrium in which both strategies are played with equal probability. Most subjects preferred to play the game rather than flipping a coin. Obviously, the expected probability of successful coordination can be larger than 50%, such

that even if risk aversion is stronger under strategic uncertainty than under pure randomization, subjects may still opt for playing the coordination game. In a slightly different game, called the matching pennies game, the same authors found a preference for the coin flip, however, replicating a finding in Camerer and Karjalainen (1994). In this game, there exists only a mixed equilibrium within which both the strategies are played with equal probability: this is the case because for each event where player 1 wins money, player 2 loses the same amount, and vice versa. Fox and Weber (2002) also found that the perceived competence of the opponent/partner affected the choice between the game and the risky lottery. A similar result is reported in Eichberger et al. (2008) for more complex games played against opponents with a different perceived degree of rationality. These findings imply that the perception of the game structure and the perception of the opponent strongly influence attitudes toward strategic uncertainty, showing the importance of subtle social influences.

The situation of strategic uncertainty is common to other games as well. Two-player examples include the ultimatum game and the trust game, also known as investment games. In the ultimatum game, an endowment is given to one player, called the proposer, whose task it is to propose a division of the endowment between herself and the second player, called the responder. Given this proposal, the responder has to make a binary decision between accepting and rejecting the proposal. In the former case, each player receives her share as specified in the proposal. In the latter case, both players receive nothing. Assuming pure self-interest, any positive amount is preferable to a zero payoff, and the subgame perfect Nash equilibrium therefore predicts that the responder accepts any positive amount. Hence, the proposer proposes the smallest positive unit to the responder, keeping the rest of the endowment to herself. If players have distributional or reciprocal preferences, however, proposers may offer significant amounts to the responder. Similarly, responders may reject small offers because of such preferences. The proposer obtains a risky decision situation because of the uncertainty regarding social preferences of the responder, and thus the possibility of rejection and loss of the endowment. Bellemare et al. (2008) show that proposers perceive the situation as a risky decision problem, and make proposals according to their beliefs about the probability of a rejection by the responder for each potential proposal of shares. This implies that proposers will offer more in situations where the responder is able to reject offers as described above, compared with situations where the responder is forced to accept any proposal (i.e., a dictator game). This prediction is empirically confirmed by Bellemare et al. (2008, see also Bolton and Zwick 1995).

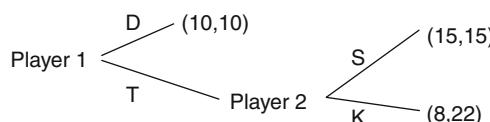
The above evidence from coordination games that a risk-averse person may be even less inclined to take social risk than an equivalent lottery risk has been tested intensively in the trust game. In the trust game, the first player receives an endowment  $e$  and can submit any part  $x$  of it to the second player, the trustee. The remaining amount  $(e - x)$  she keeps for herself. For any amount  $x$  that is entrusted to her, the trustee then receives an amount  $b \times x$  from the experimenter, with  $b$  typically set equal to 2 or 3 in experiments. The real-world equivalent for the multiplication factor  $b$  is that the trustee may have access to more profitable investment opportunities than the first player, but she does not have enough funds to make the investment with her own money. After receiving  $b \times x$ , the trustee in turn decides how much to send back to the first player, denoted  $y$ . The final payoffs for the first and the second player are, respectively,  $e - x + y$  and  $(b \times x) - y$ . The game is either played with  $x$  and  $y$  to be freely picked by the players, or in a binary form with only two possible amounts for  $x$  and  $y$ , respectively. In the latter case, for example, the first player can submit either the full

endowment or zero, and the second player may either keep the full amount  $b \times x$  or send back the equal split ( $b \times x/2$ ).

The decision to trust by player 1 in the trust game is a decision implying social risk, and the relevant payoff probabilities depend on her beliefs about the trustee's behavior. A variety of studies tried to identify whether first players in the trust game indeed perceive the game as a simple decision under risk, or whether there is an additional, social component. Using behavioral measures of trust game choices and risk-decision behavior, Eckel and Wilson (2004) find no correlation between various measures of risk attitude from lottery choices, and the trust game behavior. Schechter (2007) finds evidence of a positive correlation between trusting and risk seeking, using a risky decision task that matches the structure of the trust game as closely as possible. Schechter's results show that the relation between trusting and risk seeking is far from one to one: one unit invested in a risky lottery increases the amount submitted in the trust game by 0.28 units. Kosfeld et al. (2005) study the trust game from a biological perspective. They administer either the peptide oxytocin or a placebo to first players in the trust game. They find that oxytocin increases trust behavior, but does not affect risk-seeking behavior in a matched risky decision lottery task. Their result thus shows that trust includes an additional social risk component that is not related to risk seeking in uncertainty deriving from nature, and is uniquely affected by the administered hormone.

Bohnet and Zeckhauser (2004) and Bohnet et al. (2008) devise a clever experimental design to clearly separate the pure risk aspect from the additional social component using behavioral data. They call the latter component *betrayal aversion*, and show that it is substantial and that it exists across a wide range of cultural groups. Bohnet and Zeckhauser separate betrayal aversion from risk aversion by the following task. Given a certain payoff structure in the trust game (see Fig. 22.3), they ask player 1 to indicate the lowest acceptable probability of a randomly matched trustee playing equal split (instead of keeping most of the pie) such that she (player 1) would choose trusting. Given the same game structure, in a control group, player 2's choice is replaced by a random lottery draw with unknown probability. Here, they ask first movers what the lowest probability of the equal split is that would make them choose the risky strategy.

Clearly, the more risk averse a subject is, the higher will be the lowest acceptable probability of the good outcome. If trust is simply determined by risk attitude, this probability should be equal in both questions. In fact, Bohnet and coauthors find that people require much higher probabilities in the trust game than in the pure risk game to make them accept the more risky trust option. The difference between a typical minimum acceptable probability in the trust



**Fig. 22.3**

**Trust game used to identify betrayal aversion.** Notes: The game is played sequentially. Player 1 either distrusts (D) or trusts (T). Conditional on strategy T by player 1, player 2 either shares equally (S), or betrays the trust, keeping (K) most of the pie for herself.  $(x,y)$  indicate a payoff of  $x$  for player 1 and  $y$  for player 2

game and in the risk game measures betrayal aversion. The authors report probability differences in the range of 10–20% points more requested in the trust game, indicating a much stronger aversion to social risk than to natural risk from an impartial random device.

In sum, there is clear evidence that people perceive strategic uncertainty as a risky decision problem. Attitudes toward this uncertainty deviate strongly from individual choice setting, however, and depend on the social content of the game. There is some indication that risk attitudes and beliefs are correlated, amplifying risk aversion in social contexts.

## Further Research

---

The previous sections have explored recent research efforts in economics that bring social factors to the forefront of decision making under risk. While we have seen many interesting findings that increase the realism and descriptive power of traditional theories of decision making under risk and uncertainty, one element that has become apparent is the many contradictions between different studies. Since effects may often be subtle and this literature is still in its infancy, this is not surprising: as the research field becomes more mature, more of these contradictions will be explained, and the discovery of the underlying causes will greatly improve our understanding of social influences in risk taking. Beyond resolving the underlying controversies in the baseline research, there is also a wealth of applications to real-world phenomena to be explored. Since such real-world applications may be more interesting – and are generally the drivers behind the baseline research we have already discussed – the present section will pay heed to the latter in order to pinpoint lacunae that remain in our understanding of social influences on risk attitudes. We will thereby focus on two areas that are currently highly debated – financial agency and climate change.

Principal–agent relationships in financial markets constitute an especially promising terrain for future research. As discussed briefly earlier, such relationships can be complex, involving potentially different payoff mechanisms for the agent and the principal. A better empirical understanding of such asymmetric compensation mechanisms seems fundamental for the design of agency contracts, in which the principal tries to induce an agent to represent her interests. Influencing the agent's risk attitude in such a way as to be aligned with the interests of the principal can be quite challenging and the mechanisms needed to achieve this are still poorly understood, as shown by the results of Lefebvre and Vieider (2010) discussed above. Many other – even more involved – relationships exist in the real world, which we are only starting to study and understand. One important question that has emerged from the recent financial turmoil is, for instance, how the principal–agent relationships can be reformed in cases where agents are employed by one principal (e.g., a bank or a financial advisor), but take decisions for or provide advice to another principal (e.g., a private investor). Indeed, incentives of such agents to act exclusively in the interest of the permanent employer have been blamed for advice that has often been excessively risk-laden from the point of view of clients, and which is at least partially to blame for the escalating risk spiral culminating in the 2008 financial crisis. While models and theories on how to fix such relationships do exist, it seems crucial to obtain empirical evidence to test such models and investigate subtleties that may have been missed in the deductive process of their construction.

The imitation of others around us is one of the most natural and common ways of learning about the world and improving oneself. This imitation may however take very different forms.

One could learn from someone else how to better perform a given task – or one could learn how not to perform a task from observing somebody else's success or failure (Offerman and Schotter 2009 and references therein). Since actual decision-making processes are rarely observable in sufficient detail, a focus on the outcomes of such decision-making processes can generally be considered a more accessible substitute. Problems may arise from this approach, however, when outcomes are largely dependent on chance, and observers are more likely to observe successful decision makers than the unsuccessful ones. This is sometimes known as a *survivor bias*, whereby only successful actors may stay in the market long enough to be observed, while the unsuccessful ones tend to drop out and disappear from sight relatively quickly. When survival depends largely on chance, imitating the behavior of survivors may not lead to better decision-making processes, and may sometimes lead to worse decisions. A typical example of this may be bubble formation in financial markets. When markets get artificially inflated, observers may imitate the behavior of early movers and buy increasing amounts of stocks, thus pushing their price even higher and thus increasing the risk of a painful burst. Unfortunately, the social element in these types of behavior is still poorly understood, such that it is unclear what part of such behavior is due to uninformed following, and which part is instead caused by rational speculation (as long as this self-enforcing mechanism continues, the price will continue going up, even though the asset is already overpriced).

Whenever the linkage between actions and outcomes is poorly understood, delegating decisions involving risk or uncertainty creates complex issues of incentives and accountability of agents. This derives at least in part from the fact that in an uncertain world it is often difficult to determine what outcomes result from skill and what outcomes result from pure luck (Maboussin 2010), with consequent difficulties for the establishment of effective incentive and responsibility mechanisms. Moreover, perceptions by the principal of the control that agents have over outcomes in different contexts and the subsequent preferences for accountability regimes resulting from such perceptions may depend crucially on the principal's ideological worldview (Tetlock and Vieider 2010). Bartling et al. (2009) gave principals a choice between a low-trust strategy leaving little room for employee discretion and a high-trust strategy that gave employees substantial freedom, but also provided the possibility of substantial rent sharing. They found that the trust strategy resulted in substantial benefits for both employers and employees when there was competition between employees who could build up reputations. Notwithstanding this clear advantage, some employers never switched to the trust strategy, thus sacrificing potential payoffs. These employers may have been deterred by the risk that employees may turn out to be untrustworthy and shirk.

While the latter examples show how important social perceptions and mechanisms can be, they also provide an indication of how poorly understood some of the mechanisms involved still are. Reputation formation, different accountability regimes, reporting requirements, etc., all play into the equation and interact with the pure incentive mechanisms on which economic theory has classically focused. Expanding the traditional focus and allowing for more complex interactions, which are generally present in the real world, promise to improve our understanding of the existing processes as well as opening up new directions for contract design. In this sense, integrating the variety of findings from different social science literatures seems a promising direction for moving forward. Huge parallel literatures on principal-agent relationships exist in economics and organizational science, and even sociology and organizational and social psychology. To mention, but one example, organizational scientists have long called for abandoning the simplistic assumption of invariant risk aversion by agents still central to the

agency literature in economics (Wiseman and Gomez-Mejia 1998) – an assumption whose anachronism has been dramatically unveiled by recent events. Unfortunately, such communication between different disciplines in the social sciences is still rather the exception than the rule, be it because of the different taxonomies or methodologies adopted that make their integration arduous work, or because of turf wars between scientific communities driven by narrow self-interest.

Financial markets are by no means the only real-world phenomenon on which a better understanding of social aspects of decision making under uncertainty promises to shed some light. Let us take another issue that currently occupies a prominent position in international policy agendas: climate change. It is often repeated how humanity's failure to control emissions of greenhouse gases constitutes a classic tragedy of the commons problem (Hardin 1968). Such problems have been intensively studied in economics under the label of public good games, and the large body of evidence clearly rejects the economically selfish prediction of zero contributions to the common good (Andreoni 1995). Nevertheless, this large body of knowledge only offers poor guidance for the complex phenomenon of emission reductions, for the simple reason that the issue differs from the classic public good problem on several dimensions that may be expected to be relevant for the decision-making process.

First of all, climate change is a public bad rather than a public good, and there is solid evidence that different frames significantly influence people's decisions. Perhaps more importantly from the point of view of the present chapter, the public bad of climate change – or alternatively, the public good of emissions reductions – involves risks that do not derive only from the possible failure of other group members to contribute. Mounting levels of greenhouse gas concentrations in the atmosphere increase global average temperatures. In turn, such mounting temperatures produce changes in global weather patterns, resulting in increased likelihoods of catastrophic events such as prolonged droughts, floods, increased storm frequency and intensity, and even potential migrations and wars resulting from such events. Finally, the likelihoods of different events themselves are often poorly understood because of previously unexperienced feedback cycles such as the recession of the permafrost zones in northern Siberia, etc.

All this shows the point that we are here considering a risky public good involving unknown and highly uncertain probabilities – a type of phenomenon on which virtually no evidence exists (Zeckhauser 2006). To this, one needs to add the social connotations of the problem that go far beyond the classical social issues involved in public good games as studied in the experimental economics literature. The asymmetric wealth of the actors involved (with some of the poorer actors also being the largest emitters of greenhouse gases; see Reuben and Riedl (2009) for the effects of such asymmetries on public good contributions), the fact that some of the largest emitters are unlikely to face the highest risks (Sunstein 2008), different historical responsibilities for stocks of greenhouse gases present in the atmosphere, and different cultural backgrounds all constitute social elements that complicate the process, and which are still often poorly understood in isolation, let alone combined. For instance, how may risk attitudes – and hence the willingness to reduce existing or future risks – depend on income or wealth of the involved actors? And how does this dependency interact with historical responsibility for current risks by the richest actors, combined with a power for stopping increases in risk level that rests mainly with relatively poor actors? Obtaining better evidence on the drivers of decision-making processes in situations that have complex social and risk dimensions appears to be crucial for a better understanding of involved decision-making processes and international coordination on these complex issues.

While the examples provided here are by no means exhaustive, they seem apt at providing a snapshot of how more evidence on the social drivers of decision-making processes under risk can help us better understand – and thus potentially improve – complex real-world interactions. The challenge will be to dissect these complex issues in such a way as to be able to draw causal inferences on single aspects in isolation, as well as to later reassemble the collected evidence in order to understand their interaction. One discipline or research methodology will hardly be able to rise to this challenge. Classical laboratory experiments from economics aimed at understanding basic decision-making processes will need to be integrated with evidence from more complex decision environments in the field. A better understanding of social interactions from social psychology will be just as important as a solid underpinning from organizational science, and deeper insights into the legal aspects necessary for binding international agreements and their implementation into national law. Finally, philosophy can come into the picture by highlighting the intercultural common ground of ethical thought and shared principles on which the indispensable closer international cooperation can be founded. Paraphrasing a famous scientist from another discipline: everything should be made as simple as possible, but not simpler.

## References

- Aldashev G, Kirchsteiger G, Sebald A (2010) How (not) to decide: procedural games. Discussion paper, Ecares Brussels
- Andreoni J (1995) Cooperation in public goods experiments: kindness or confusion? *Am Econ Rev* 85:891–904
- Anscombe FJ, Aumann RJ (1963) A definition of subjective probability. *Ann Math Stat* 34:199–205
- Arkes HR, Dawes RM, Christensen C (1986) Factors influencing the use of a decision rule in a probabilistic task. *Organ Behav Hum Decis Process* 37:93–110
- Asch S (1955) Opinions and social pressure. *Sci Am* 193:31–35
- Barber BM, Odean T (2001) The internet and the investor. *J Econ Perspect* 15:41–54
- Barber BM, Odean T (2002) Online investors: do the slow die first? *Rev Financ Stud* 15:455–487
- Barber BM, Heath C, Odean T (2003) Good reasons to sell: reason-based choice among group and individual investors in the stock market. *Manage Sci* 49:1636–1652
- Bartling B, Fehr E, Schmidt K (2009) Screening, competition, and job design: economic origins of good jobs. Working paper, University of Munich
- Bateman I, Munro A (2005) An experiment on risky choice amongst households. *Econ J* 115:C176–C189
- Bell DE (1982) Regret in decision making under uncertainty. *Oper Res* 30:961–981
- Bellemare C, Kröger S, Van Soest A (2008) Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. *Econometrica* 76:815–839
- Bernheim DB (1994) A theory of conformity. *J Polit Econ* 102:841–877
- Bernoulli D (1954/1738) Exposition of a new theory on the measurement of risk. *Econometrica* 22:23–36
- Bohnet I, Frey BS (1999) The sound of silence in prisoner's dilemma and dictator games. *J Econ Behav Organ* 38:43–57
- Bohnet I, Zeckhauser R (2004) Trust, risk and betrayal. *J Econ Behav Organ* 55:467–484
- Bohnet I, Greig F, Herrmann B, Zeckhauser R (2008) Betrayal aversion – evidence from Brazil, China, Switzerland, Turkey, the United Arab Emirates, and the United States. *Am Econ Rev* 98:294–310
- Boles TL, Messick DM (1995) A reverse outcome bias: the influence of multiple reference points on the evaluation of outcomes and decisions. *Organ Behav Hum Decis Process* 61:262–275
- Bolton G E, Ockenfels A (2009) Risk taking and social comparison. A comment. *Am Econ Rev* 100:628–633
- Bolton GE, Zwick R (1995) Anonymity versus punishment in ultimatum bargaining. *Game Econ Behav* 10:95–121
- Bolton GE, Brandts J, Ockenfels A (2005) Fair procedures: evidence from games involving lotteries. *Econ J* 115:1054–1076

- Bond CF, Titus LJ (1983) Social facilitation: a meta-study of 241 studies. *Psychol Bull* 94:265–292
- Borah A (2010) Other-regarding preferences and procedural concerns. Discussion paper, University of Pennsylvania
- Bossaerts P (2009) What decision neuroscience teaches us about financial decision making. *Ann Rev Financ Econ* 1:383–404
- Camerer CF, Karjalainen R (1994) Ambiguity aversion and non-additive beliefs in noncooperative games: experimental evidence. In: Munier B, Machina MJ (eds) Models and experiments in risk and rationality. Kluwer, Dordrecht, pp 325–358
- Cappelen AW, Konow J, Sorensen EO, Tungodden B (2009) Just luck: an experimental study of risk taking and fairness. Discussion paper, Bergen University Business School
- Chakravarty S, Harrison G, Haruvy EE, Rutström EE (2010) Are you risk averse over other people's money? *South Econ J* 77:901–913
- Charness G, Jackson MO (2009) The role of responsibility in strategic risk taking. *J Econ Behav Organ* 69:241–247
- Cialdini RB (1993) Influence: the psychology of persuasion. William Morrow, New York
- Conradt L, List C (2010) Group decisions in humans and animals: a survey. *Philos Trans R Soc B* 364:719–742
- Cooper D, Rege M (2008) Social interaction effects and choice under uncertainty: an experimental study. Discussion paper, Florida State University
- Corazzini L, Greiner B (2007) Herding, social preferences, and (non-)conformity. *Econ Lett* 97:74–80
- Curley SP, Yates JF, Abrams RA (1986) Psychological sources of ambiguity avoidance. *Organ Behav Hum Decis Process* 38:230–256
- De Palma A, Picard N, Ziegelmeyer A (2010) Individual and couple decision behavior under risk: evidence on the dynamics of power balance. *Theory Decis* (forthcoming)
- Delgado MR, Schotter A, Ozbay EY, Phelps EA (2008) Understanding overbidding: using the neural circuitry of reward to design economic auctions. *Science* 321:1849–1852
- Diamond PA (1967) Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility: comment. *J Polit Econ* 75:765–766
- Diecidue E, van de Ven J (2008) Aspiration level, probability of success and failure, and expected utility. *Int Econ Rev* 49:683–700
- Duflo E, Saez E (2002) Participation and investment decisions in a retirement plan: the influence of colleagues' choices. *J Public Econ* 85:121–148
- Dufwenberg M, Muren A (2006) Generosity, anonymity, gender. *J Econ Behav Organ* 61:42–49
- Eckel CC, Wilson RK (2004) Is trust a risky decision? *J Econ Behav Organ* 55:447–465
- Eichberger J, Kelsey D, Schipper BC (2008) Granny versus game theorist: ambiguity in experimental games. *Theory Decis* 64:333–362
- Ellsberg D (1961) Risk, ambiguity, and the savage axioms. *Q J Econ* 75:643–669
- Engelmann JB, Capra M, Noussair C, Berns GS (2009) Expert financial advice neurobiologically "offloads" financial decision-making under risk. *PLoS ONE* 4:e4957
- Eriksen KW, Kvaloy O (2010) Myopic investment management. *Rev Financ* 14:521–542
- Falk A, Ichino A (2006) Clean evidence on peer effects. *J Labor Econ* 24:39–57
- Fox CR, Tversky A (1995) Ambiguity aversion and comparative ignorance. *Q J Econ* 110:585–603
- Fox CR, Weber M (2002) Ambiguity aversion, comparative ignorance, and decision context. *Organ Behav Hum Decis Process* 88:476–498
- Frisch D, Baron J (1988) Ambiguity and rationality. *J Behav Decis Mak* 1:149–157
- Gneezy U, Potters J (1997) An experiment on risk taking and evaluation periods. *Q J Econ* 112:631–645
- Goeree JK, Yariv L (2006) Conformity in the lab. Discussion paper, Caltech
- Gong M, Baron J, Kunreuther H (2010) Group cooperation under uncertainty. *J Risk Uncertain* 39:251–270
- Haigh MS, List JA (2005) Do professional traders exhibit myopic loss aversion? Experimental analysis. *J Finance* 60(1):523–534
- Haisley E, Weber RA (2010) Self-serving interpretations of ambiguity in other-regarding behavior. *Game Econ Behav* 68:614–625
- Haisley E, Mostafa R, Loewenstein GF (2008) Subjective relative income and lottery ticket purchases. *J Behav Decis Mak* 21:283–295
- Hardin G (1968) The tragedy of the commons. *Science* 162:1243–1248
- Harsanyi JC (1955) Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *J Polit Econ* 63:309–321
- Heinemann F, Nagel R, Ockenfels P (2009) Measuring strategic uncertainty in coordination games. *Rev Econ Stud* 76:181–221
- Herstein IN, Milnor J (1953) An axiomatic approach to measurable utility. *Econometrica* 21:291–297
- Hoffman E, McKabe K, Smith VL (1996) Social distance and other-regarding behavior in dictator games. *Am Econ Rev* 86:653–660
- Huck S, Weizsäcker G (1999) Risk, complexity, and deviations from expected-value maximization: results of a lottery choice experiment. *J Econ Psychol* 20:699–715

- Inbar Y, Cone J, Gilovich T (2010) People's intuitions about intuitive insight and intuitive choice. *J Pers Soc Psychol* 99:232–247
- Isenberg DJ (1986) Group polarization: a critical review and meta-analysis. *J Pers Soc Psychol* 50:1141–1151
- Kahneman D, Miller DT (1986) Norm theory: comparing reality to its alternatives. *Psychol Rev* 93:136–153
- Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. *Econometrica* 47:263–291
- Knight F (1921) Risk, uncertainty, and profit. University of Chicago Press, Chicago
- Kocher MG, Trautmann ST (2010) Selection into auctions for risky and ambiguous prospects. *Econ Inq* (forthcoming)
- Konana P, Balasubramanian S (2005) The social-economic-psychological model of technology adoption and usage: an application to online investing. *Decis Support Syst* 39:505–524
- Kosfeld M, Heinrichs M, Zak PJ, Fischbacher U, Fehr E (2005) Oxytocin increases trust in humans. *Nature* 435:673–676
- Krawczyk M (2011) A model of procedural and distributive fairness. *Theory Decis* 70:111–128
- Krawczyk M, Le Lec F (2010) "Give me a chance!" An experiment in social decision under risk. *Exp Econ* 13:500–511
- Kroll Y, Davidovitz L (2003) Inequality aversion versus risk aversion. *Economica* 70:19–29
- Kruglanski AW, Freund T (1983) The freezing and unfreezing of lay-inferences: effects on impressional primacy, ethnic stereotyping, and numerical anchoring. *J Exp Soc Psychol* 19:448–468
- Latane B (1981) The psychology of social impact. *Am Psychol* 36:343–356
- Lefebvre M, Vieider FM (2010) Reigning in excessive risk taking by executives: experimental evidence. GATE Working Paper No. 1006, University of Lyon
- Lerner JS, Tetlock PE (1999) Accounting for the effects of accountability. *Psychol Bull* 125:255–275
- Linde J, Sonnemans J (2009) Social comparison and risky choices. Discussion paper, University of Amsterdam
- Loomes G, Sugden R (1982) Regret theory: an alternative theory of rational choice under uncertainty. *Econ J* 92:805–824
- Maboussin MJ (2010) Untangling skill and luck. How to think about outcomes – past, present, and future. Legg Mason Capital Management strategy paper
- Machina MJ (1989) Dynamic consistency and non-expected utility models of choice under uncertainty. *J Econ Lit* 27:1622–1668
- Miller PM, Fagley NS (1991) The effects of framing, problem variation, and providing rationale on choice. *Pers Soc Psychol Bull* 17:517–522
- Muthukrishnan AV, Wathieu L, Xu AJ (2009) Ambiguity aversion and persistent preference for established brands. *Manage Sci* 55:1933–1941
- Offerman T, Schotter A (2009) Imitation and luck: an experimental study on social sampling. *Game Econ Behav* 65:461–502
- Rabin M, Weizsäcker G (2009) Narrow bracketing and dominated choices. *Am Econ Rev* 99:1508–1543
- Read D, Loewenstein GF, Rabin M (1999) Choice bracketing. *J Risk Uncertain* 19:171–197
- Reuben E, Riedl A (2009) Public goods provision and sanctioning in privileged groups. *J Confl Resolut* 53:72–93
- Reynolds DB, Joseph J, Sherwood R (2009) Risky shift versus cautious shift: determining differences in risk taking between private and public management decision-making. *J Bus Econ Res* 7:63–78
- Rohde I, Rohde K (2009) Risk attitudes in a social context. Discussion paper, Erasmus University
- Sacerdote B (2001) Peer effect with random assignment: results for Dartmouth roommates. *Q J Econ* 116:681–704
- Savage LJ (1954) The foundations of statistics. Wiley, New York
- Schechter L (2007) Traditional trust measurement and the risk confound: an experiment in rural Paraguay. *J Econ Behav Organ* 62:272–292
- Schmidt U, Starmer C, Sugden RF (2008) Third-generation prospect theory. *J Risk Uncertain* 36:203–223
- Sebald A (2010) Attribution and reciprocity. *Game Econ Behav* 68:339–352
- Shafir E, Simonson I, Tversky A (1993) Reason-based choice. *Cognition* 49:11–36
- Sieck W, Yates JF (1997) Exposition effects on decision making: choice and confidence in choice. *Organ Behav Hum Decis Process* 70:207–219
- Stoner JAF (1961) A comparison of individual and group decisions under risk. Master's thesis, Massachusetts Institute of Technology (unpublished)
- Sugden R (2003) Reference-dependent subjective expected utility. *J Econ Theory* 111:172–191
- Sunstein CR (2008) The world vs. the United States and China? The complex climate change incentives of the leading greenhouse gas emitters. *UCLA Law Rev* 55:1675–1700
- Sutter M (2009) Individual behavior and group membership: comment. *Am Econ Rev* 99:2247–2257
- Sutter M, Czermak S, Feri F (2010) Strategic sophistication of individuals and teams in experimental normal-form games. IZA discussion paper 4732
- Takemura K (1993) The effect of decision frame and decision justification on risky choice. *Jpn Psychol Res* 35:36–40

- Takemura K (1994) Influence of elaboration on the framing of decision. *J Psychol* 128:33–40
- Taylor K (1995) Testing credit and blame attributions as explanation for choices under ambiguity. *Organ Behav Hum Decis Process* 64:128–137
- Tetlock PE, Boettger R (1994) Accountability amplifies the status quo effect when change creates victims. *J Behav Decis Mak* 7:1–23
- Tetlock PE, Vieider FM (2010) Ideology, agency and accountability: explaining shifting managerial preferences for alternative accountability regimes. Working paper, University of California, Berkeley
- Trautmann ST (2009) A tractable model of process fairness under risk. *J Econ Psychol* 30:803–813
- Trautmann ST (2010) Individual fairness in Harsanyi's utilitarianism: operationalizing all-inclusive utility. *Theory Decis* 68:405–415
- Trautmann ST, Wakker PP (2010) Process fairness and dynamic consistency. *Econ Lett* 109(3):187–189
- Trautmann ST, Zeckhauser R (2010) Blindness to the benefits of ambiguity: the neglect of learning opportunities. Discussion paper, Harvard University
- Trautmann ST, Vieider FM, Wakker PP (2008) Causes of ambiguity aversion: known versus unknown preferences. *J Risk Uncertain* 36:225–243
- Van Huyck JB, Battalio RC, Beil RO (1990) Tacit coordination in games, strategic uncertainty, and coordination failure. *Am Econ Rev* 80:234–248
- Vieider FM (2009) The effect of accountability on loss aversion. *Acta Psychol* 132:96–101
- Vieider FM (2011) Separating real incentives and accountability. *Experim Econ* (forthcoming)
- Von Neumann J, Morgenstern O (1947) Theory of games and economic behavior. Princeton University Press, Princeton
- Wallach MA, Kogan N, Bem DJ (1964) Diffusion of responsibility and level of risk taking in groups. *J Abnorm Soc Psychol* 68:263–274
- Weigold MF, Schlenker BR (1991) Accountability and risk taking. *Pers Soc Psychol Bull* 17:25–29
- Wiseman RM, Gomez-Mejia LR (1998) A behavioral agency model of managerial risk taking. *Acad Manage Rev* 23(1):133–153
- Zajonc RB (1965) Social facilitation. *Science* 149:269–274
- Zeckhauser R (2006) Investing in the unknown and unknowable. *Cap Soc* 1(2):1–39
- Zizzo DJ (2004) Inequality and procedural fairness in a money-burning and stealing experiment. *Res Econ Inequal* 11:215–247

## Risk Perception



# 23 Risk Intelligence

Dylan Evans

University College Cork, Cork, Ireland

<i>Introduction</i> .....	604
<i>Objections to My Definition of Risk Intelligence</i> .....	605
<i>Measuring Risk Intelligence</i> .....	607
<i>The Distribution of Risk Intelligence in the General Population</i> .....	608
<i>The Dunning–Kruger Effect</i> .....	613
<i>Why Is Risk Intelligence Important?</i> .....	614
<i>Methods for Increasing Risk Intelligence</i> .....	616
<i>Further Research</i> .....	617
<i>Appendix</i> .....	618

**Abstract:** Risk intelligence is the ability to estimate probabilities accurately. In this context, accuracy does not imply the existence of objective probabilities; on the contrary, risk intelligence presupposes a subjective interpretation of probability. Risk intelligence can be measured by calibration testing. This involves collecting many probability estimates of statements whose correct answer is known or will shortly be known to the experimenter, and plotting the proportion of correct answers against the subjective estimates. Between 1960 and 1980, psychologists measured the calibration of many specific groups, such as medics and weather forecasters, but did not gather extensive data on the calibration of the general public. This chapter presents new data from calibration tests of over 6,000 people of all ages and from a wide variety of countries. High risk intelligence is rare. Fifty years of research in the psychology of judgment and decision-making shows that most people are not very good at thinking clearly about risky choices. They often disregard probability entirely, and even when they do take probability into account, they make many errors when estimating it. However, there are some groups of people with unusually high levels of risk intelligence. Lessons can be drawn from these groups to develop new tools to enhance risk intelligence in others. First, such tools should accustom users to specifying probability estimates in numerical terms. Second, they should focus on a relatively narrow area of expertise, if possible. Thirdly, these tools should provide the user with prompt and well-defined feedback. Regular calibration testing might fulfill all three of these requirements, though training assessors by giving them feedback about their calibration has shown mixed results. More research is needed before we can reach a definitive verdict on the value of this method.

## Introduction

---

Although the term “risk intelligence” has been gaining currency during the past few years, there is still no consensus as to what it means. According to David Apgar, the term denotes “an individual’s or an organization’s ability to weigh risks effectively,” and involves “classifying, characterizing, calculating threats; perceiving relationships; learning quickly; storing, retrieving, and acting upon relevant information; communicating effectively; and adjusting to new circumstances” (Apgar 2006). According to Frederick Funston, coauthor of *Surviving and Thriving in Uncertainty: Creating the Risk Intelligent Enterprise* (Funston and Wagner 2010), risk intelligence is “the ability to effectively distinguish between two types of risks: the risks that must be avoided to survive by preventing loss or harm; and, the risks that must be taken to thrive by gaining competitive advantage,” and involves the ability to “translate these insights into superior judgment and practical action to improve resilience to adversity and improve agility to seize opportunity” (Krell 2010). Funston is a principal at Deloitte & Touche LLP, and Deloitte seems keen for people to associate the phrase “risk intelligence” with its brand, to judge by the series of research papers they have published on this topic and their release of an iPhone app which purports to let users create a “risk intelligence map.”

The trouble with both of these definitions is that they are rather vague, and encompass so many abilities as to be practically useless, and certainly immune to any kind of scientific measurement. I prefer a much more restricted definition: risk intelligence is the ability to estimate probabilities accurately (Evans 2012). Not only is this concept simpler to grasp and more precise than those suggested by Apgar and Funston, it is also more susceptible to measurement. Before explaining how to measure risk intelligence as I define it, however, I will first address some common objections I have encountered when explaining my definition.

## Objections to My Definition of Risk Intelligence

---

The most common objection to my proposed definition seems to be that it makes no reference to notions of harm, threat or danger, which some people consider to be central to the concept of risk. The observation is correct, but I regard this feature as a virtue of my approach rather than a fault. While it is true that the term “risk” is intimately associated in the vernacular with undesirable possibilities, those who study the subject largely agree that this restriction is somewhat arbitrary. Risk experts never tire of pointing out that an exclusive focus on “downside risk” tends to encourage a risk-averse attitude and to discourage an awareness of the potential rewards from exploiting risky opportunities. Indeed, Funston goes so far as to make this point central to his definition of risk intelligence, as we have already seen. My definition of risk intelligence avoids value judgments altogether by referring simply to “probabilities,” regardless of whether such probabilities refer to outcomes that we find pleasant or not.

Another objection to my definition comes from those who are fond of the distinction between “risk” and “uncertainty” proposed by the American economist Frank Knight in 1921. In that year Knight published his influential book, *Risk, Uncertainty, and Profit*, in which he argued that:

- ▶ Uncertainty must be taken in a sense radically distinct from the familiar notion of Risk, from which it has never been properly separated. The term “risk,” as loosely used in everyday speech and in economic discussion, really covers two things which, functionally at least, in their causal relations to the phenomena of economic organization, are categorically different. . . . The essential fact is that “risk” means in some cases a quantity susceptible of measurement, while at other times it is something distinctly not of this character; and there are far-reaching and crucial differences in the bearings of the phenomenon depending on which of the two is really present and operating. . . . It will appear that a measurable uncertainty, or “risk” proper, as we shall use the term, is so far different from an unmeasurable one that it is not in effect an uncertainty at all. “We . . . accordingly restrict the term ‘uncertainty’ to cases of the non-quantitative type” (Knight 1921).

This distinction has become so influential that economists now talk about “Knightian uncertainty” when referring to risks that are immeasurable or impossible to calculate, in contrast to risks that can be quantified. This way of explicating the distinction is misleading, however, since both risk and Knightian uncertainty can be measured and quantified. The putative distinction is really about the way in which the probabilities are calculated in each case.

The prototypical case of risk, in Knight’s sense, is a casino game like roulette or blackjack. To work out what the odds of a given bet should be in these games, all you need to know is the rules of the game itself. You do not need to collect any data or observe how the game is actually played. You can just read the rule book in the comfort of your armchair and work it out with pen and paper (though a laptop would often help considerably).

This may be contrasted with, say, working out what the odds of a given bet should be in a horse race. In this case, it will not help much if you simply read the rule book. In addition to this, you also need to gather lots of data about the horses in the race, and the jockeys, and the racetrack, and the likely weather on the day of the race, and who knows what else. You can get these data in all sorts of ways – reading the “form” of the horses as published in newspapers, looking carefully at the horses with your own eyes, talking to tipsters, listening to the weather forecast, and so on. And then you need to crunch all these data and come up with an estimate of how likely it is that this horse will win this race.

One way to crunch the data is to use a computer. Another is to use your brain; first absorb the data by reading, watching, and listening, and then mull them over in your own head and come up with an estimate of how likely it is that the horse will win. Doing that well is what I call risk intelligence.

I'm sticking to my guns, and will continue to refer to this ability as *risk* intelligence (rather than, say, *uncertainty* intelligence) because I do not think the distinction proposed by Frank Knight holds water. For one thing, the mere fact that the odds in a racetrack world cannot be worked out from first principles does not mean they cannot be measured or quantified. On the contrary, when gamblers or bookmakers estimate the chances of a horse winning a race or a team winning a basketball match, what else are they doing if not quantifying uncertainty?

More importantly, pure casino worlds do not exist – except in the pages of economics textbooks. In this respect, I agree with maverick trader Nassim Nicholas Taleb, who wrote in his bestselling book, *The Black Swan*, that:

- ▶ In real life you do not know the odds; you need to discover them, and the sources of uncertainty are not defined. Economists, who do not consider what was discovered by noneconomists worthwhile, draw an artificial distinction between Knightian risks (which you can compute) and Knightian uncertainty (which you cannot compute), after one Frank Knight, who rediscovered the notion of unknown uncertainty and did a lot of thinking but perhaps never took risks, or perhaps lived in the vicinity of a casino. Had he taken financial or economic risk he would have realized that these “computable” risks are largely absent from real life! They are laboratory contraptions! (Taleb 2007).

When Taleb states that computable risks are “laboratory contraptions,” he means that casino worlds are artificial entities which have to be deliberately manufactured under sterile conditions, like an unstable element that only exists for a few brief moments in a physics lab. It took thousands of years for the irregular-shaped knucklebones used in ancient Rome, and the Vibhīdaka nuts used in ancient India, to evolve into the precision dice used in modern casinos, with their pips drilled and then filled flush with a paint of the same density as the acetate, such that the six numbers are equally probable. It takes even greater engineering prowess to produce a fair roulette wheel. The manufacturers of roulette wheels perform elaborate tests to ensure that the numbers generated are truly random, and even then the wheels still have flaws, allowing some cunning players to make a fortune with a “biased wheel attack.” In 1873, for example, a mechanic from Lancashire called Joseph Jagger identified a biased wheel in Monte Carlo and won the equivalent of \$70,000 in one day.

Taleb tells a lovely story to illustrate the unreality of casino worlds. A casino in Las Vegas thought it had all the bases covered in risk management. It was sufficiently diversified across the various tables to not have to worry about taking a hit from lucky gamblers. It had a state-of-the-art surveillance system to catch cheats. But the four largest risks faced by the casino in the past few years lay completely outside its risk management framework. For example, it lost around \$100 million when an irreplaceable performer in the main show was maimed by his tiger. In other words, even casinos are not pure casino worlds.

A third objection to my proposed definition is that it makes no reference to the concept of risk appetite. Again, I think this is an advantage rather than a defect. Risk intelligence is a cognitive capacity, a purely intellectual ability to estimate probabilities accurately. It involves gauging the extent of one's knowledge on a given topic, and can be objectively assessed to distinguish between those with high risk intelligence and those with low RQ. Risk *appetite*, on the other hand, is an emotional trait. It has to do with preferences. Some people enjoy taking on

risk, and other people avoid risk like the plague; some people are willing to expose themselves to danger, while others prefer to shield themselves from as many losses as possible. Unlike risk intelligence, there's no right or wrong about risk appetite; it's just a matter of taste.

## Measuring Risk Intelligence

---

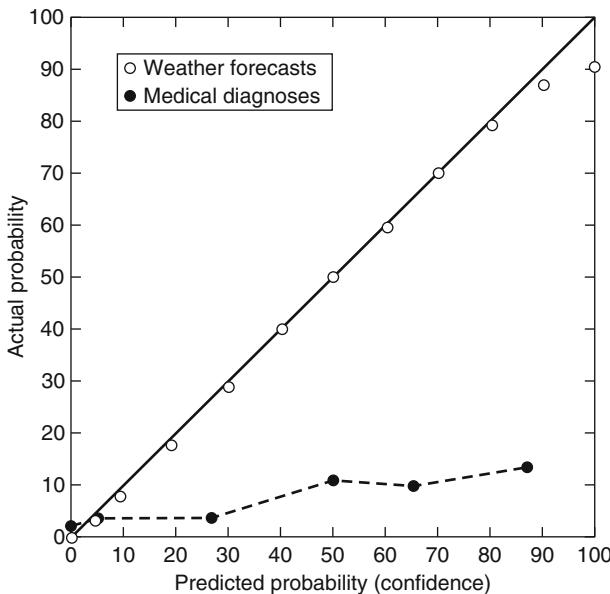
How do we judge the accuracy of probability estimates? One way is to compare subjective probability estimates to objective statistics. For example, one can ask people to estimate the probability of death from various causes for some particular demographic group, and compare these estimates to the mortality data. This method is restricted, of course, to subject areas for which data are readily available.

Another way to measure a person's ability to provide accurate probability estimates is calibration testing (Lichtenstein et al. 1982). This involves collecting many probability estimates about statements whose correct answer is known or will shortly be known to the experimenter, and plotting the proportion of correct answers against the subjective estimates. For example, suppose that every day you estimate the probability that it will rain in your neighborhood the following day, and then you note whether or not it did, in fact, rain on each day. To simplify things a little, let us assume that you can only choose from a discrete set of probability values, such as 0, 0.1, 0.2, etc. Over the course of a year, you collect 365 estimates, for each of which you have also indicated whether it did, in fact, rain or not. Suppose that you estimated the chance of rain as 0 on 15 days. If you are well calibrated, it should have rained on none of those days. Again, if there were 20 days which you assigned a 0.1 probability of rainfall, it will have rained on 2 of those days if you are well calibrated. Perfect calibration, in other words, would be shown by all points falling on the identity line.

The nice thing about this method is that, unlike comparing subjective estimates to objective statistics, it does not take objective probabilities to be conceptually prior. Throughout this chapter, I use numerical probabilities to express degrees of belief – or, to put it another way, to quantify subjective uncertainty. For the sake of fairness, however, I should point out that this is a minority view. A rival school of thought holds that numerical probabilities refer to objective facts about the world, namely, long run frequencies. According to this view, the statement that “there is a 50% chance of this coin landing on heads” does not have anything to do with anyone’s beliefs; rather, it means that, in the long run, the coin will land heads up on half of all the times it is tossed.

These two schools of thought are happy to use the same mathematical tools; the probability calculus is uncontroversial. They differ only in their interpretation of what the math means. The subjectivist school is the older one; when Jacob Bernoulli first showed how any probability could be represented as a number, it was degrees of belief that he had in mind. During the twentieth century, however, the frequentist approach became more popular, and this is now the dominant view. For reasons I do not have time to go into here, I think the frequentist view is fundamentally flawed.

According to the subjectivist view, there is no such thing as a “true” probability, in the sense of some objective fact existing out there in the world; probabilities are just numerical expressions of our subjective degree of belief. I suppose a subjectivist could say that a probability estimate is “true” when it accurately expresses the strength of one’s conviction – indeed, that is a vital aspect of risk intelligence – but that is a far cry from the frequentist view of probabilities as facts.

**Fig. 23.1**

Calibration curves for US weather forecasters and for doctors. (Data are from Murphy and Winkler 1977 and Christensen-Szalanski and Bushyhead 1981)

Calibration tests might be seen as incorporating something of the frequentist approach, in the sense that they plot probability estimates (subjective probabilities) against the proportion of correct predictions (an objective measure). But the proportion of correct predictions is not the same thing an “objective probability.” So while I sometimes talk loosely of “making accurate probability estimates,” strictly speaking this phrase is incoherent. The accuracy of an estimate can only be measured by comparing it to some objective fact, and such facts do not exist in the case of probabilities. This is why experts who study risk intelligence usually prefer to speak of “well-calibrated” probability estimates rather than of accurate ones.

Nobody is perfectly calibrated, but as you can see from [Fig. 23.1](#), US weather forecasters are pretty close. However, as the same figure also shows, doctors are very badly calibrated.

The doctors whose data are shown in [Fig. 23.1](#) were asked estimate the probability that real patients had pneumonia after taking a medical history and completing a physical examination. When these doctors estimated that there was about a 5% chance that a patient had pneumonia, they were about right. But only about 15% of the patients to whom these doctors assigned a 90% chance of pneumonia turned out to have the disease. In other words, these doctors had much more faith in the accuracy of their diagnoses than was justified by the evidence.

## The Distribution of Risk Intelligence in the General Population

Between 1960 and 1980, psychologists measured the calibration of many specific groups, such as medics (Christensen-Szalanski and Bushyhead 1981) and weather forecasters (Murphy and

Winkler 1977), but did not gather extensive data on the calibration of the general public (the data from these two groups are shown in Fig. 23.1, above). In their survey of the research up to 1980, Lichtenstein and colleagues report studies of hospital patients, psychologists, military personnel, engineering students, and other groups, but no large cross-sectional studies of the general population (Lichtenstein et al. 1982). One reason for this was no doubt because the testing was done with pen and paper, which made data collection and processing a time-consuming process. It appears that interest in calibration testing began to decline after 1980, and has not progressed much since then. This area of research is ripe for revival, especially now that the Internet allows testing and data collection to be automated.

In this section I present data from calibration tests of over 6,000 people of all ages and from a wide variety of countries. I was able to collect such a large amount of data by using an online calibration test rather than a pencil-and-paper version.

In December 2009 my co-investigator Benjamin Jakobus and I created an online calibration test (<http://www.projectionpoint.com>), and promoted the site through press releases, media interviews, blogs, and Internet discussion forums. The test consisted of 50 statements (see Appendix), below each of which were 11 buttons indicating percentage values ranging from 0 to 100 in increments of ten. Visitors to the site were instructed to indicate how likely they thought it was that each statement was true according to the following rules:

- If you are absolutely sure that a statement is true, you should click on the button marked 100%.
- If you are completely convinced that a statement is false, you should click on the button marked 0%.
- If you have no idea at all whether it is true or false, you should click on the button marked 50%.
- If you are fairly sure that it is true, but you are not completely sure, you should click on 60%, or 70%, or 80%, or 90%, depending on how sure you are.
- If you are fairly sure that it is false, but you are not completely sure, you should click on 40%, or 30%, or 20%, or 10%, depending on how sure you are.

After participants had answered all 50 questions in the test, they were asked if they would like to take part in our study. If they declined, they were given their test results, and then their data were deleted from the server. If they agreed, they were asked to specify the following demographic details: gender, nationality, age, highest level of academic education, and profession. They were then given their test results.

The test results were calculated as follows. First, all the times that the participant assigned a likelihood of 0% to a statement were counted, and then we counted how many of those statements were actually true. We proceeded in the same way for each of the other likelihoods and plotted each data point on a graph with the probability estimates on the  $x$ -axis and the proportion of correct answers in each category on the  $y$ -axis. We always plotted the points for the categories of 0% and 100% (if a participant never used the 100% category, we plotted that as 100% correct), but we only plotted points for the other categories if the participant had used them at least three times. We then connected the points by a continuous line. This line is henceforth referred to as the participant's calibration curve.

As already noted, a perfect calibration curve would lie on the identity line  $x = y$ . The further away from that diagonal line the curve lies, the poorer the calibration is. We created a simple index of a participant's calibration by calculating the area between their calibration curve and

the identity line and scaling the result to a number between 0 and 100, where 100 = perfect calibration (i.e., this number is inversely proportional to the size of the area between the calibration curve and the identity line). Henceforth we refer to this number as the “RQ score.”

We chose to use this measure, rather than using the better-known Brier score (Brier 1950) for three reasons. Firstly, our approach is much easier to understand for a lay audience. Secondly, we find some of the statistical properties of the Brier score to be unsatisfactory. Suppose a person chooses  $p_i$  for  $n_i$  questions, of which  $k_i$  are correct. The Brier score would yield  $(p_i - k_i/n_i)^2$ , which implies that it is irrelevant whether  $k_i = 1$  and  $n_i = 2$  or  $k_i = 1,000$  and  $n_i = 2,000$ . We object to this because we believe that the scores should be weighted by the number of answers. We could remedy this by adjusting the Brier score as follows:  $n_i \times (p_i - k_i/n_i)^2$  (Aaron Brown, personal communication), but again we prefer a simpler approach. Finally, the Brier score is a composite measure of calibration, resolution, and knowledge (Murphy 1973), whereas we wish to measure only calibration.

Some people realize fairly quickly that there is an easy way to game this test. If a participant always selects the 50% category – and if the test contains equal numbers of true and false statements – they will obtain an RQ score of 100. The high score generated by this strategy does not reflect good calibration, however, so to remedy this we created a second indicator which we call the “K factor” (for Keynes and Keats). To calculate the K factor, each time that a participant uses the categories 10%, 20%, 30%, 40%, 60%, 70%, 80%, or 90%, they get one point. When they use 0%, 50%, or 100% they get zero. The maximum K factor is therefore 50 for a fifty-question test. The K factor gives an indication of how reliable a participant’s RQ score is as an indicator of their calibration.

Between December 10, 2009, and February 6, 2010, a total of 21,910 people visited the web site, of whom 10,187 took the online calibration test and gave us permission to use their data in our research. We removed all participants with an RQ score of 0 ( $N = 30$ ) on the grounds that this was probably due to error. We then removed all those with a K score of less than 10 ( $N = 3,295$ ) on the grounds that in such cases the RQ score would not be a good indicator of calibration. We then removed all those who did not specify their gender ( $N = 154$ ) or their educational achievement ( $N = 10$ ). After these adjustments, a total of 6,698 participants remained in our sample.  [Table 23.1](#) shows the composition of this sample by gender and educational achievement.

The participants’ ages ranged from under 10 to over 80, though most participants were aged either 21–30 ( $N = 2,242$ ) or 31–40 ( $N = 1,852$ ). Participants came from every continent,

 **Table 23.1**  
**Composition of the sample by gender and educational achievement**

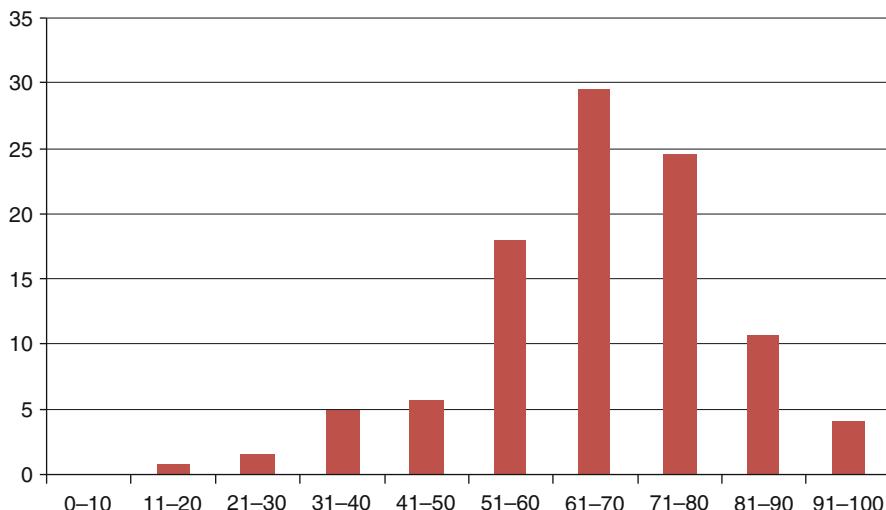
Education	Men	Women	Total
Primary or less	37	10	47
Secondary	974	223	1,197
First degree	2,617	675	3,292
Masters	1,113	338	1,451
Ph.D.	567	144	711
Total	5,308	1,390	6,698

with over 20 nationalities, with the most well-represented countries being (in order): the USA ( $N = 2,633$ ), the UK ( $N = 1,118$ ), Ireland ( $N = 1,023$ ), Canada ( $N = 402$ ), Australia ( $N = 343$ ), and Germany ( $N = 188$ ).

The mean RRQ score for the sample of 6,698 was 65.02. ► *Figure 23.2* shows the distribution of RQ scores in this sample.

As shown in ► *Table 23.2*, the mean RQ score of the men in our sample was significantly higher than the mean RQ score of the women (two-tailed  $t$ -test for independent samples, not assuming equal variance,  $p < .0001$ ; similar results were found with a Mann–Whitney rank sum test). When analyzed according to educational achievement, however, the difference between the mean RQ scores of the men and women in our sample is only significant in those whose highest level of educational achievement is a first degree or masters. As the graph in ► *Fig. 23.3* makes clearer, education seems to make little or no difference to the calibration of women until they achieve a Ph.D., while every increase in educational achievement seems to improve calibration in men. It is this “education effect” that explains the higher mean RQ score of men in our sample, since it contains a high proportion of people whose highest level of educational achievement is a first degree or masters (70.7%). It is only at this level of achievement that the education effect produces a clear difference in calibration between men and women. Something about university education seems to boost calibration in men to levels significantly higher than those in women, and this gap only closes when people have attained the highest level of educational achievement – the Ph.D.

Previous studies have not found differences in calibration between men and women (Lichtenstein and Fischhoff 1981) or between people with different levels of education (Lichtenstein and Fischhoff 1977). However, this may be due to the fact that most studies have typically involved fewer than 200 participants, and participants have generally been required to provide a smaller number of probability estimates – both of which have severely

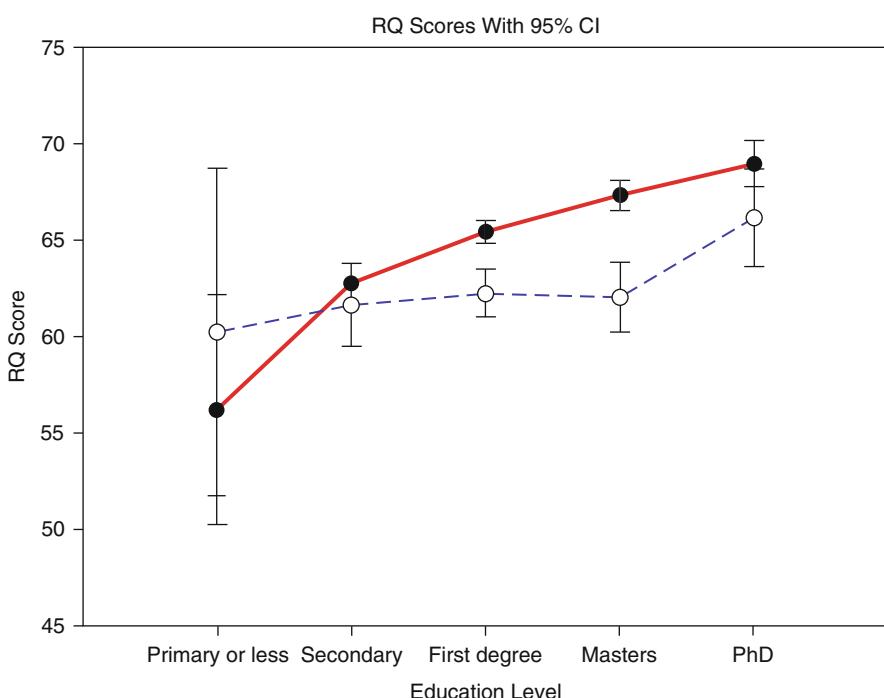


► **Fig. 23.2**

The distribution of RQ scores in the sample of 6,698 people reported here. The x-axis shows the RQ score, and the y-axis the percentage of the sample in each category

**Table 23.2**Mean RQ scores by gender and educational achievement (\*=significant at  $p < .05$ )

Education	Men		Women		Standard error difference	95% CI of the difference		<i>p</i>
	Mean RQ	(Stdev)	Mean RQ	(Stdev)		Lower	Upper	
	Mean RQ	(Stdev)	Mean RQ	(Stdev)		Lower	Upper	
Primary or less	56	18.55	60	13.71	5.30	-15.12	7.07	0.427
Secondary	63	16.25	62	16.58	1.21	-1.26	3.49	0.365
First degree	65	15.20	62	16.03	0.66	1.87	4.47	<.0001*
Masters	67	13.50	62	16.67	0.99	3.32	7.22	<.0001*
Ph.D.	69	14.21	66	15.30	1.41	0.05	5.60	0.046*
ALL	66	15.11	63	16.22	0.48	2.21	4.10	<.0001*

**Fig. 23.3**

Effect of education on calibration in men and women (dashed line=women)

limited the capacity of previous research to detect individual differences with a high enough degree of statistical significance. Our sample of 6,698 people, in which each participant provided 50 probability estimates, provides a dataset which is an order of magnitude larger than any previous study, and this may have permitted us to detect patterns which were previously invisible.

## The Dunning–Kruger Effect

---

Calibration tests measure the extent to which one is able to gauge how much one knows, not knowledge per se. The extent of one's knowledge should, therefore, make no difference to one's score on a calibration test, and empirical research seems to bear this out (Lichtenstein and Fischhoff 1977). If education improves calibration in men, as it seems to, this effect must therefore be due to something other than the greater knowledge that greater education typically bestows. A possible candidate for this other factor is enhanced metacognition. Metacognition refers to “the ability to know how well one is performing, when one is likely to be accurate in judgment, and when one is likely to be in error” (Kruger and Dunning 1999). Kruger and Dunning (1999) have argued that the skills that engender competence in a domain are often the very same skills necessary to *evaluate* competence – including one's own competence – in that domain. Education may therefore make people not only more knowledgeable, but also more aware of the limits of their knowledge. Contrary to what Lichtenstein and Fischhoff (1977) concluded, then, it does appear that “those who know more also know more about how much they know,” but as Fig. 23.3 shows, this dictum seems to apply more to men than to women. Conversely, as Charles Darwin noted, “ignorance more frequently begets confidence than does knowledge” (Darwin 1871).

If education has this effect in men, why does it not also have this effect in women? I confess that I am at a loss to explain this apparent sex difference. In the rest of this section, I will limit myself to some general remarks about the Dunning–Kruger effect and risk intelligence in both sexes.

Like the Roman god Janus, risk intelligence has two faces looking in opposite directions. One looks outward at the external world, and attempts to gather objective data that will throw light on the matter at hand. The other looks inward and attempts to assess how much relevant knowledge one really has. Good probability estimates require that both faces see clearly – in other words, risk intelligence requires both objective and subjective knowledge (cognition and metacognition).

To change the metaphor, picture your mind as a light bulb shining in an otherwise darkened room. Some nearby objects are fully illuminated; you can see them in every detail, present and identifiable. These are the things you know very well – the names of your friends, what you had for breakfast this morning, how many sides a triangle has, and so on. The objects on the other side of the room are completely shrouded in darkness. These are the things about which you know nothing – the five thousandth digit of pi, the composition of dark matter, or King Nebuchadnezzar's favorite color. Between the light and the darkness, however, lies a gray area in which the level of illumination gradually shades away. In this twilight zone, the objects are not fully illuminated, but neither are they completely invisible. You know something about these things, but your knowledge is patchy and incomplete – the law of the land (unless you are a lawyer), the evidence for climate change (unless you are a climatologist), and the causes of the

credit crunch (even economists are still arguing about this). Risk intelligence involves gauging exactly how illuminated the objects in this twilight zone really are.

Nothing in our education system or our culture prepares us to operate in the twilight zone. If we are cautious, we relegate everything beyond the zone of complete illumination to complete obscurity, not daring to venture an opinion on things of which we do, in fact, have some inkling. If we are overconfident, we do the opposite, expressing views about things in the twilight zone with more conviction than is justified. It is hard to steer between these two extremes; daring to speculate, but with prudence. Yet that is what risk intelligence is all about.

This is why calibration tests which require users to estimate the likelihood of general knowledge statements are a perfectly acceptable way to measure risk intelligence. When putting a probability value on these statements, one is required to weigh up all the relevant evidence that one possesses and gauge one's true level of uncertainty on the matter. There is no reason to restrict the content of these statements to threats, dangers, and other concepts generally associated with the vernacular view of risk.

Nor should it matter if a calibration test is dominated by questions that refer to a particular area of knowledge or a particular part of the world. A number of people in the study reported in section ➤ [The Distribution of Risk Intelligence in the General Population](#) complained that they thought the test had a “US bias.” However, since the test measures how well one is able to gauge how much one knows, rather than knowledge per se, any such bias is irrelevant.

Of course, if a calibration test contained too many questions about which a user knew absolutely nothing, and which the user should therefore use the 50% category many times, the result would not be an accurate measure of the person’s level of risk intelligence. To provide a good measure of risk intelligence, it is necessary to gather probability estimates across the whole range of possible values from 0% to 100%. This is what the K-factor described in section ➤ [The Distribution of Risk Intelligence in the General Population](#) was intended to capture.

## Why Is Risk Intelligence Important?

---

The doctors whose data are graphed in ➤ [Fig. 23.1](#) were extremely overconfident in their diagnoses. When a patient had a 15% chance of having pneumonia, they would give them a 90% chance. That meant they were likely to recommend more tests than were strictly necessary, prescribe more treatments than were warranted, and cause their patients needless worry.

It is not always true to say “better safe than sorry,” either. Some tests are invasive and painful, and there are many cases where the treatment prescribed for an ailment that is not present can be harmful to the patient. It is always better to make a choice on the basis of accurate information than the basis of error, no matter what the context. And the data clearly show that these diagnoses are full of error.

Doctors are not the only professionals who require good risk intelligence. Finance professionals are also required to estimate probabilities as a regular part of their work. Yet as the recent financial crisis has shown us, bankers and those who work at credit rating agencies can make basic errors when assessing the likelihood of certain outcomes. What would happen if trading desks implemented some kind of regular calibration testing, or if a requirement for calibration training were incorporated into new financial regulations? The Basel Committee on Banking Supervision, which formulates broad supervisory standards for financial

institutions around the world, is currently working on a new update to the Basel Accords. Basel II required numerical assessments of both the probability of default and the expected loss given default, and it specifically forbade relying on rating agency assessments to estimate these, but these estimates were typically produced by computer models. What if future rounds of the Basel Accords were to include some provision for testing risk intelligence? The head of a prop desk in a bank, for example, could ask each trader to estimate the probability that each trade will make a profit and keep track of each trader's calibration.

The commitment, detection, and investigation of crime also offer many opportunities for the exercise of risk intelligence. For example, when police officers question suspects, they must judge whether the responses they receive are truthful or not. It is rarely the case that such judgments are clear cut; more typically, the officer has some index of suspicion which lies somewhere between complete confidence and absolute distrust. In other words, the question of whether a suspect is lying usually demands a probability estimate rather than a simple yes or no. As with all probability estimates, their accuracy may be measured by means of a calibration test when the true answers are known. Psychologists have carried out many such tests, and the results all point to the same conclusion; police officers and other professional investigators are all massively overconfident about their ability to discern lies. Although they are convinced they can spot deception, their real ability to sift fact from fiction is scarcely better than flipping a coin. One reason for this is that most police officers look at the wrong signals; shifty eyes, for example, are not a good signal of deception, and yet it is one that many investigators rely on. This has serious consequences for the criminal justice system, and law enforcement officials should therefore be trained in risk intelligence if we are to reduce miscarriages of justice.

High levels of risk intelligence will also be required among the general population if we are to deal effectively with any of the big challenges that humanity faces in the twenty-first century. Climate change is a case in point. Nobody knows precisely how increasing levels of greenhouse gasses in the atmosphere will affect the climate in different regions around the globe. The Intergovernmental Panel on Climate Change (IPCC) does not make definite predictions; instead, it sets out a variety of possible scenarios and attaches different probabilities to them to indicate the level of uncertainty associated with each one. Knowing how to make sense of this information is crucial if we are to allocate resources sensibly to the various alternative solutions, from carbon trading schemes to the development of alternative energy sources, or even planetary-scale geo-engineering. How can citizens make informed decisions about such matters if they are not equipped to think clearly about risk and uncertainty? Too often, the public figures who take opposite views about climate change make exaggerated claims which convey greater certainty than is warranted by the evidence. Critics dismiss the claims of the IPCC out of hand, while believers in climate change proselytize with equal dogmatism. Both kinds of exaggeration seriously hamper informed debate; the latter kind also terrifies kids. One survey of 500 American preteens found that one in three children between the ages of 6 and 11 feared that the earth would not exist when they reached adulthood because of global warming and other environmental threats (Lomborg 2009). We see the same pattern in the UK, where a survey showed that half of young children aged between seven and eleven are anxious about the effects of global warming, often losing sleep because of their concern. Without the tools to understand the uncertainty surrounding the future of our climate, we are left with a choice between two equally stupid alternatives – ignorant bliss or fearful paralysis.

## Methods for Increasing Risk Intelligence

---

Leading medical schools around the world are beginning to wake up to the problem of low risk intelligence among doctors. Something called “confidence-based assessment” is increasingly being used in these schools. In this form of assessment, students must not only give the right answer, but also assess the confidence with which they give each answer. If a student gives the wrong answer confidently, that receives the worst possible grade; if they give the wrong answer but are not confident, then they get a better grade; giving the right answer but without confidence is OK, but not ideal, as in reality it could end up with them wasting time having to consult others; and the best answer is that which is correct and made with confidence. This form of assessment is intended to help students know when to consult others (or text books etc) and when to act independently.

All well and good, but how do you prepare students for this kind of test? Nothing in our current repertoire of educational tools and methods seems well designed to equip someone with the skills to be confident when justified but doubtful when necessary. In medical schools in particular, doubt has often been perceived as a sign of weakness.

The same could be said of the financial sector, where similar macho attitudes played no small part in stoking the bubble that burst in late 2007. What if the bankers who were making all those dubious loans in the preceding decade had undergone regular calibration testing, of the sort described here? It is an interesting thought.

The fact that weather forecasters are so much better calibrated than doctors suggests that one's level of risk intelligence is not relatively fixed like IQ, but susceptible to improvement given the right conditions. Sarah Lichtenstein, an expert in the field of calibration testing, speculates that several conditions favor the weather forecasters (Lichtenstein et al. 1982). First, they have been expressing their forecasts in terms of numerical probability estimates for many years; since 1965, US National Weather forecasters have been required to say not just whether or not it will rain the next day, but how likely they think this is in actual percentage terms. They have got used to putting numbers on such things, and as a result are better at it. Doctors, on the other hand, are under no such obligations. They remain free to be as vague as they like.

Second, the task for weather forecasters is repetitive. The question to be answered (“Will it rain?”) is always the same. Doctors, however, must consider all sorts of different questions every day: “Does he have a broken rib?” “Is this growth malignant?” “How will she respond to a different type of antidepressant?”

Finally, the feedback for weather forecasters is well defined and promptly received. This is not always true for doctors. Patients may not come back, or may be referred elsewhere. Diagnoses may remain uncertain. Most theories of learning emphasize the need for rapid feedback; the longer the delay between an action (or, in this case, a prediction) and a corrective signal, the lower the chance that the later information will enable the recipient to profit from it.

These speculations could assist the development of tools to enhance risk intelligence. First, such tools should accustom users to specifying probability estimates in numerical terms. Second, they should focus on a relatively narrow area of expertise, if possible. Thirdly, these tools should provide the user with prompt and well-defined feedback. Regular calibration testing might fulfill all three of these requirements, though training assessors by giving them feedback about their calibration has shown mixed results. It should be pointed out however, that only a few studies have been carried out in this area, and they are now several decades old. More research is needed before we can reach a definitive verdict on the value of this method.

Another approach to improving calibration involves requiring people to think of reasons why they might be wrong. In one study, subjects took two calibration tests similar to the one described in section [The Distribution of Risk Intelligence in the General Population](#). In the second test, one group was asked to write down a reason supporting each of their answers, another group was asked to write down a reason contradicting each answer, and a third group wrote down two reasons, one supporting and one contradicting. Only the group asked to write down contradicting reasons showed improved calibration in the second test. This suggests that one partial remedy for overconfidence is to search for reasons why one might be wrong (Koriat et al. 1980).

## Further Research

---

As already noted, it appears that interest in calibration testing began to decline after 1980, and research in this area has not progressed much since then. I believe this area is ripe for revival, especially now that the Internet allows testing and data collection to be automated. The study reported in section [The Distribution of Risk Intelligence in the General Population](#) is ongoing, and a further 25,000 people have taken the online calibration test between the date when the dataset reported there was collected and the present time of writing (April 2011), bringing the total sample size to over 35,000 participants so far. Analysis of this growing dataset is ongoing, and further features are being added to the calibration test, including the measurement of time taken to complete the task.

In a study of horse handicappers, Steven Ceci and Jeffrey Liker found that handicapping expertise had zero correlation with IQ (Ceci and Liker 1986). IQ is the best single measure of intelligence that psychologists have, because it correlates with so many cognitive capacities. Indeed, it is this very correlation that underpins the concept of “general intelligence.” The discovery that expertise in handicapping does not correlate at all with IQ means that, whatever cognitive capacities are involved in estimating the odds of a horse winning a race may be, they are not a part of general intelligence. Or, to put it the other way round, IQ is unrelated to some real-world forms of cognitive complexity that are clear-cut cases of intelligence.

Of course, not everyone is comfortable with blanket terms like “general intelligence.” One of its most high-profile critics, the psychologist Howard Gardner, prefers to conceive of intelligence as consisting of multiple, special-purpose skill sets (Gardner 1983). It was Ceci and Liker’s paper that first led me to wonder if the ability to estimate probabilities accurately, and make wise decisions under uncertainty, might constitute a special kind of intelligence to be added to Gardner’s list.

Gardner identifies eight different kinds of intelligence: bodily-kinesthetic, interpersonal, verbal-linguistic, logical-mathematical, naturalistic, intrapersonal, visual-spatial, and musical. None of these involves an ability to estimate probabilities accurately, and yet the study of Ceci and Liker shows that this is a complex cognitive skill that some people are very good at, and this suggests that it could constitute a ninth kind of intelligence that Gardner had not considered.

To clinch this argument, I would have to show that risk intelligence is typically implemented in the brain by a specific set of neural pathways, for Gardner’s framework restricts intelligences to cognitive capacities that can be localized neurologically. Linguistic intelligence, for example, is rooted in certain structures in the brain’s left hemisphere. I believe that recent

neuroscience imaging studies would also support a distinct neural architecture for risk assessment, but further research is needed to confirm this.

High risk intelligence is rare. Fifty years of research in the psychology of judgment and decision-making shows that most people are not very good at thinking clearly about risky choices. They often disregard probability entirely, and even when they do take probability into account, they make many errors when estimating it. Like most psychologists, I had assumed that these patterns of bias were universal – until I read the paper by Ceci and Liker. These two young psychologists seemed to have stumbled on a rare breed of individuals who had somehow escaped the influence of the well-known cognitive biases that affect most peoples' ability to judge risk – expert gamblers.

As has already been noted, it appears that US weather forecasters form another group with unusually high levels of risk intelligence (Murphy and Winkler 1977). Further research might identify yet more such groups. For example, finance professionals are probably too diverse and heterogeneous to constitute such a group, but there may be particular kinds of finance professional who display higher than average risk intelligence. I suspect that hedge fund principals may be one such subgroup.

Risk intelligence differs utterly from what we normally consider intelligence to be – which is why, when we get it wrong – when banks fail, doctors misdiagnose, and weapons of mass destruction turn out not to exist in a country we have invaded – we are in such a bad position to understand the reasons. We need gamblers and weather forecasters – because we can learn an enormous amount from them, not just about money and rainfall, but about the way we make decisions in all aspects of our lives.

## Appendix

---

The 50 true/false statements in the online calibration test were as follows:

A one followed by 100 zeros is a Googol	T
Africa is the largest continent	F
Alzheimer's accounts for under half the cases of dementia in the USA	F
An improper fraction is always less than 1	F
Armenia shares a common border with Russia	F
There have been over 40 US Presidents	T
In 1994, Bill Clinton was accused of sexual harassment by a woman called Paula Jones	T
Canberra is the capital of Australia	T
Cats are not mentioned in the Bible	T
Christianity became the official religion of the Roman empire in the third century AD	F
Commodore Matthew Perry compelled the opening of Japan to the West with the Convention of Kanagawa in 1870	F
El Salvador does not have a coastline on the Caribbean	T
Gout is known as "the royal disease"	F
Harry Potter and the Goblet of Fire tells the story of Harry Potter's third year at Hogwarts	F

Humphrey Bogart had two wives before Lauren Bacall	F
In 2008 the population of Beijing was over 20 million people	F
In the Old Testament, Jezebel's husband was Ahab, King of Israel	T
Iron accounts for over 30% of the Earth's composition	T
It is possible to lead a cow upstairs but not downstairs, because a cow's knees cannot bend properly to walk back down	T
Lehman Brothers went bankrupt in September 2008	T
LL Cool J got his name from the observation "Ladies Love Cool James"	T
Male gymnasts refer to the pommel horse as "the pig"	T
Mao Zedong declared the founding of the People's Republic of China in 1949	T
More than 10 American states let citizens smoke marijuana for medical reasons	T
More than 8 out of 10 victims infected by the Ebola virus will die in 2 days	T
Most of the terrorists who carried out the attacks on 9/11 were from Saudi Arabia	T
Mozart composed over 1,000 works	F
Natural gas has an odor	F
Of all Arab nations, Lebanon has the highest percentage of Christians	T
Over 40% of all deaths from natural disasters from 1945 to 1986 were caused by earthquakes	T
Over 50% of Nigeria's population lives on less than \$1 per day	T
Stalagmites grow down, and stalactites grow up	F
The Italian musical term adagio means that the music should be played quickly	F
The Euphrates river runs through Baghdad	F
The face on a \$100,000 bill is that of Woodrow Wilson	T
The Islamic Resistance Movement is better known to Palestinians as Hizbollah	F
The Japanese were largely responsible for building most of the early railways in the US West	F
The last Inca emperor was Montezuma	F
The most frequently diagnosed cancer in men is prostate cancer	T
The only stringed symphonic instrument that has a pedestal and a crown is a double bass	F
The president of Russia is Vladimir Putin	F
The San Andreas Fault forms the tectonic boundary between the Pacific Plate and the North American Plate.	T
The US civil war broke out the same year the federal government first printed paper money	T
The US Declaration of Independence begins: "We the People of the United States..."	F
The word "robot" was coined by the American science fiction writer, Isaac Asimov	F
The world's highest island mountain is Mauna Kea	T
The Taj Mahal was built by Emperor Shah Jahan in memory of his favorite wife	T
There are more people in the world than chickens	F
There are no diamond fields in South America	F
Wikipedia was launched in 1999 by Jimmy Wales and Larry Sanger	F
Number of True Statements	25

## References

---

- Apgar D (2006) Risk intelligence: learning to manage what we don't know. Harvard Business School Press, Cambridge, MA
- Brier GW (1950) Verification of forecasts expressed in terms of probability. *Mon Weather Rev* 78(1):1–3
- Ceci SJ, Liker IK (1986) A day at the races: s study of IQ, expertise, and cognitive complexity. *J Exp Psychol Gen* 115:255–266
- Christensen-Szalanski JJJ, Bushyhead JB (1981) Physicians' use of probabilistic information in a real clinical setting. *J Exp Psychol Hum Percept Perform* 7:928–935
- Darwin C (1871) The descent of man. John Murray, London
- Evans D (2012) Risk intelligence: how to live with uncertainty. New York, Free Press
- Funston F, Wagner S (2010) Surviving and thriving in uncertainty: creating the risk intelligent enterprise. Wiley, Hoboken
- Gardner H (1983) Frames of mind: the theory of multiple intelligences. Basic Books, New York
- Knight FH (1921) Risk, uncertainty and profit. University of Michigan Library, Michigan, 2009 edition
- Koriat A, Lichtenstein S et al (1980) Reasons for confidence. *J Exp Psychol Hum Learn Mem* 6:107–118
- Krell E (2010) RiskChat: what is risk intelligence? <http://businessfinancemag.com/article/riskchat-what-risk-intelligence-0621>. Accessed 7 July 2010
- Kruger J, Dunning D (1999) Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Pers Soc Psychol* 77(6):1121–1134
- Lichtenstein S, Fischhoff B (1977) Do those who know more also know more about how much they know? *Organ Behav Hum Perform* 20:159–183
- Lichtenstein S, Fischhoff B (1981) The effects of gender and instructions on calibration. Decision research report. Decision Research, Eugene, OR, 81
- Lichtenstein S, Fischhoff B et al (1982) Calibration of probabilities: the state of the art to 1980. In: Kahneman D, Slovic P, Tversky A (eds) Judgement under uncertainty: heuristics and biases. Cambridge University Press, Cambridge, pp 306–334
- Lomborg B (2009) Scared silly over climate change. Guardian online. <http://www.guardian.co.uk/commentisfree/cif-green/2009/jun/15/climate-change-children>. Accessed 3 March 2011
- Murphy AH (1973) A new vector partition of the probability score. *J Appl Meteorol* 12:595–600
- Murphy AH, Winkler RL (1977) Reliability of subjective probability forecasts of precipitation and temperature. *J R Stat Soc C Appl Stat* 26(1):41–47
- Taleb NN (2007) The black swan: the impact of the highly improbable. Allen Lane, London

# 24 Risk Communication in Health

Nicolai Bodemer · Wolfgang Gaissmaier

Max Planck Institute for Human Development, Berlin, Germany

<b>Introduction .....</b>	<b>623</b>
<b>What Constitutes “Good” Risk Communication? .....</b>	<b>624</b>
<b>Current Practice in Health Risk Communication .....</b>	<b>625</b>
The Seven Sins in Health Care .....	627
Biased Reporting in the Medical Literature .....	627
Biased Reporting in the Media and Pamphlets .....	627
Consequences of Biased Reporting .....	629
<b>How Good Are Experts and Laypeople at Dealing with Risks and Uncertainties? .....</b>	<b>629</b>
Statistical (Il-)literacy in Health .....	630
The Concept of Numeracy .....	631
Measuring Numeracy .....	631
Numeracy in Experts and Laypeople .....	632
Consequences of Innumeracy .....	632
<b>The Role of Numbers and Words in Risk Communication .....</b>	<b>633</b>
Narrative Versus Statistical Evidence .....	633
Expressing Probabilities with Words Versus Numbers .....	635
Verbal and Numerical Probabilities in Health .....	635
Preferences for Verbal Versus Numerical Probabilities .....	636
<b>Transparent Risk Communication: How to Overcome Statistical Illiteracy and Innumeracy .....</b>	<b>637</b>
Relative Versus Absolute Risks .....	637
Conditional Probabilities Versus Natural Frequencies .....	639
Single-Event Probabilities Versus Frequencies .....	641
Five-Year Survival Rates Versus Mortality Rates .....	641
Lead-Time Bias .....	642
Overdiagnosis Bias .....	642
Graphical Representations .....	645
The Example of Icon Arrays .....	645
Graph Literacy .....	647
Uncertainty Communication .....	647
Fear of Disclosing Uncertainty .....	648
How to Communicate Uncertainty .....	648

<b>Further Research .....</b>	<b>650</b>
Research Gaps .....	650
Individual Differences .....	650
Integrating Information Sources .....	651
Implementing Theories of Risk Communication .....	651
Obstacles to Implementing Risk Communication .....	652
Teaching Statistical Literacy .....	653
Statistical Teaching in Schools .....	653
Statistics Training Education for Health Professionals .....	653
Statistics Education for (Science) Journalists .....	654
<b>Conclusion .....</b>	<b>654</b>

**Abstract:** Policy makers, health professionals, and patients have to understand health statistics to make informed medical decisions. However, health messages often follow a persuasive rather than an informative approach and undermine the idea of informed decision making. The current practice of health risk communication is often biased: Risks are communicated one sided and in nontransparent formats. Thereby, patients are misinformed and misled. Despite the fact that the public is often described as lacking basic statistical literacy skills, statistics can be presented in a way that facilitates understanding. In this chapter, we discuss how transparent risk communication can contribute to informed patients and how transparency can be achieved. Transparency requires formats that are easy to understand and present the facts objectively. For instance, using statistical evidence instead of narrative evidence helps patients to better assess and evaluate risks. Similarly, verbal probability estimates (e.g., “probable,” “rare”) usually result in incorrect interpretations of the underlying risk in contrast to numerical probability estimates (e.g., “20%,” “0.1”). Furthermore, we will explain and discuss four formats – relative risks, conditional probabilities, 5-year survival rates, and single-event probabilities – that often confuse people, and propose alternative formats – absolute risks, natural frequencies, annual mortality rates, and frequency statements – that increase transparency. Although research about graphs is still in its infancy, we discuss graphical visualizations as a promising tool to overcome low statistical literacy. A further challenge in risk communication is the communication of uncertainty. Evidence about medical treatments is often limited and conflicting, and the question arises how health professionals and laypeople deal with uncertainty. Finally, we propose further research to implement the concepts of transparency in risk communication.

## Introduction

---

Understanding health statistics is one basic prerequisite for making health decisions. Policy makers evaluate health statistics when implementing health programs, insurance companies assess the cost-effectiveness of health interventions, and doctors and patients need to know the chances of harms and benefits of different treatment alternatives. The channels available to inform decision makers about risks are manifold, and so are the ways risks can be framed. A widespread phenomenon is what we call biased reporting in risk communication. By biased, we mean two things: First, information is incomplete and one sided. For instance, benefits of a health treatment are reported, while drawbacks are omitted. Second, the continuous use of nontransparent and incomprehensible risk communication formats misleads decision makers. In this chapter, we discuss the interaction between the fact that most people have difficulties with statistical information and the way health risks can be represented. The chapter is organized as follows:

1. *What constitutes “good” risk communication?* We start this chapter by discussing the objective of risk communication.
2. *Current practice in health risk communication.* We describe current drawbacks in the practice of risk communication.
3. *How good are experts and laypeople at dealing with risks and uncertainties?* We present evidence about the public’s problems in adequately interpreting statistical information.
4. *The role of numbers and words in risk communication.* We discuss the role of narrative and verbal information, in comparison to statistical information.

5. *Transparent risk communication: How to overcome statistical illiteracy and innumeracy.* We present alternative formats that improve statistical comprehension, in contrast to those frequently used in practice.
6. *Further research.* We point out important directions for future research to make our society more risk literate.

## What Constitutes “Good” Risk Communication?

---

Of central importance, but at the same time the subject of much controversy, is the issue of what the goal of risk communication ought to be. To put it differently, what is the standard by which risk communication should be evaluated? There are at least two different perspectives a communicator can adopt: one is persuasive, the other informative (or educative). Despite commonalities between these two perspectives, there is an area of tension resulting from the different objectives each of the views follows. Let us first discuss persuasion.

The press release of the first European Randomized Study of Screening for Prostate Cancer (ERSPC) stated, “Screening for prostate cancer can reduce deaths by 20%. ERSPC is the world’s largest prostate cancer screening study and provides robust, independently audited evidence, for the first time, of the effect of screening on prostate cancer mortality” (Wilde 2009). This news was celebrated as a successful demonstration of the benefits of prostate-specific antigen (PSA) screening. Based on this statement, policy makers and doctors could argue for regular PSA tests, and men might express their willingness to participate in the screening program. However, the actual benefits of screening for prostate cancer with PSA tests are not as clear as they seem, as we will demonstrate later in this chapter. Similarly, health advertisements usually promote behavioral change. For example, an advertisement for screening for vascular diseases appeals with the admonition, “Don’t be a victim” (see Gigerenzer et al. 2007).

More extreme attempts to change people’s attitudes and intentions are fear appeals (for a meta-analysis, see Witte and Allen 2000). For instance, antitobacco campaigns show pictures of smokers’ lungs or mouth cancer to demonstrate the consequences of smoking. The aim of fear appeals is less to educate the public about health interventions than to promote and encourage health behavior change. Frosch et al. (2007) evaluated health advertisements and found that the vast majority of those aired on TV made emotional appeals, and only about one-fourth gave explicit information about risk factors, prevalence, and condition causes. These campaigns are not solely run by the pharmaceutical industry, but also by health authorities and health associations. The term “social marketing” has been coined to describe the application of “marketing principles and techniques to create, communicate and deliver value in order to influence target audience behaviors that benefit society (public health, safety, the environment, and community development) as well as the target audience” (Kotler and Lee 2007, p. 7).

The underlying assumption of the persuasive approach is that people’s motivation and ability to engage in health decisions is rather low in the first place and hence deviate from a “normative” standard – however that might be defined. From this point of view, the key measure for successful risk communication is behavioral change that is reflected in more favorable attitudes toward health (prevention) programs, higher intentions to participate, and finally higher attendance rates.

The alternative perspective – that is, the informative approach – begins with the assumption that people are able to take responsibility for their health and make individual and

informed health decisions. A decision per se is not right or wrong – it always depends on the patient's personal preferences, values, and needs. Some patients prefer watchful waiting to invasive treatments; others prefer rapid treatment of abnormalities. Some patients accept severe side effects of treatments if the benefit is high, while others do not. For instance, it has been reported that patients are willing to accept higher risks of severe side effects than their physicians (Heesen et al. 2010). The concept of health communication as information is related to the paradigm shift from the classic notion of a paternalistic doctor–patient relationship to one of shared decision making and informed consent – a mutual, interactive process between the doctor and the patient, who jointly make health decisions (e.g., Edwards and Elwyn 2009). With this in mind, the main evaluation principles in risk communication should be transparency and (gained) knowledge. Risk communication requires comprehensible, unbiased, and complete information to educate doctors and patients and provide a basis for shared decision making. Informed decisions require facts about etiological factors, epidemiological data, treatment benefits and side effects, uncertainties, and potential costs. Without knowing the risk of developing a particular disease, the chance that a treatment will lead to success, or the risk of side effects, neither policy makers nor doctors and patients can effectively make informed health decisions.

We consider ourselves proponents of the latter approach and argue that the major objective of risk communication should be informing and educating rather than persuading. However, we do not aim at discussing the “persuasion” approach.

An example contrasting the two different approaches in risk communication is given in Fig. 24.1. While the flyer “mammograms save lives” encourages women to participate in mammography screening and convey an illusion of certainty (“mammograms save lives – there's no doubt about it (...) Hope for a cancer-free future starts with you”), the facts box summarizes current scientific evidence and compares 2,000 women in a mammography group with 2,000 women not attending the screening.

## Current Practice in Health Risk Communication

Consider the following fictive example: An urologist offers a 57-year-old patient a PSA test – the previously mentioned screening test to detect early stages of prostate cancer. The patient, who has never heard of this test, says to his doctor, “Well, I don't really know. What do you think I should do?” The urologist hesitates and then answers, “I think you should do the test.” The patient agrees without knowing his baseline risk of prostate cancer or the benefits and harms of the PSA test. The patient trusts the doctor's recommendation and believes the doctor's decision was based on the best medical knowledge. However, this does not need to be true.

Doctors often practice what is called defensive decision making: They prescribe treatments that may not be best for their patients but that reduce their own risk of facing legal consequences. In our example, the doctor could have recommended not participating, because current scientific evidence does not show a benefit of PSA screening in the reduction of prostate cancer mortality (Djulbegovic et al. 2010; Sandblom et al. 2011). But even if the doctor did not believe in the efficacy of PSA screening, not urging the patient to have the test might cause trouble if the patient is later diagnosed with prostate cancer. Daniel Merenstein, an American urologist, informed his patient about the pros and cons of PSA screening, and the patient decided not to participate. Later, the patient developed prostate cancer and sued Merenstein, whose residency had to pay compensation of US\$1 million (see Gigerenzer and Gray 2011).

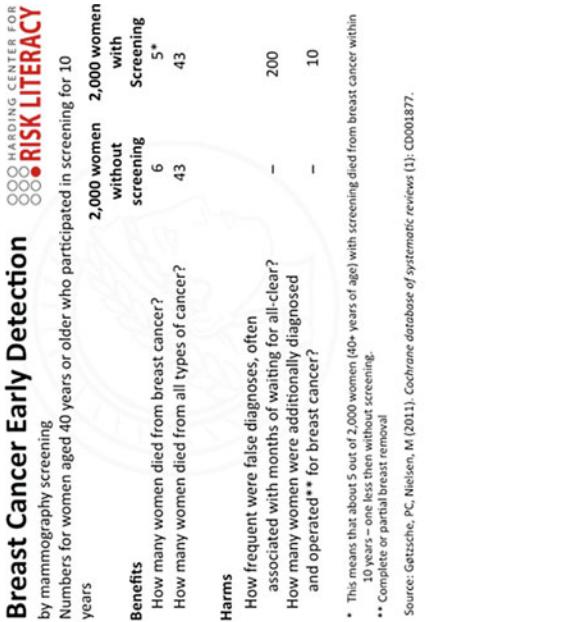
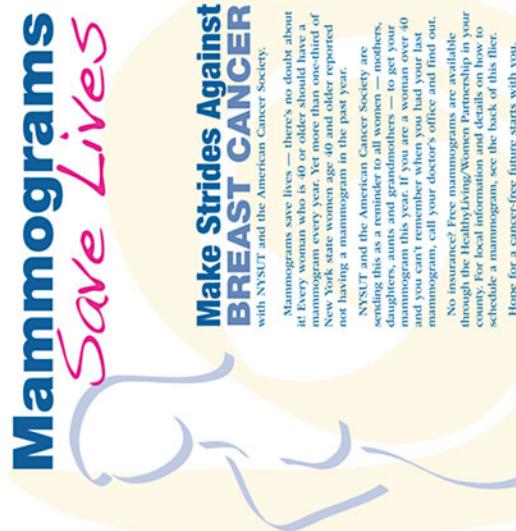


Fig. 24.1

Two different ways to inform women about mammography screening. The flier on the left side by the American Cancer Society (Retrieved from [www.nysut.org/files/makingstrides\\_070921\\_poster.pdf](http://www.nysut.org/files/makingstrides_070921_poster.pdf) in April 2011) encourages women to participate in regular mammography screening without providing information about benefits and harms of the screening program. It states that "mammograms save lives – there's no doubt about it (...) Hope for a cancer-free future starts with you." The facts box on right side (Retrieved from [www.harding-center.com/fact-boxes/](http://www.harding-center.com/fact-boxes/) in April 2011) summarizes the most important results based on the current scientific evidence and informs rather than persuades. It contrasts 2,000 women aged 40 and older who participate in mammography screening over 10 years with 2,000 of the same age who do not. Besides the benefits of the screening program, the facts box also includes information about potential harms like overtreatment

Such decisions have far-reaching consequences for doctors' behavior as well as the entire health system. For instance, many doctors in Switzerland order PSA tests for their patients but would not participate themselves (Steurer et al. 2009).

## The Seven Sins in Health Care

---

Defensive decision making is just one of seven "sins" in health care that Gigerenzer and Gray (2011) identified. They have called for the "century of the patient" to demonstrate the importance of a radical change in health policy. This change centers on fostering patients who understand health risks and who are willing to take responsibility for their own health decisions based on transparent and unbiased information. A misinformed patient is the result of the seven sins: biased funding in medical research, biased reporting in medical journals, biased reporting in pamphlets, biased reporting in the media, conflicts of interests, defensive decision making, and last but not least, doctors' lack of fundamental health literacy skills (see ➤ *Table 24.1*). Although these sins are more or less linked to each other, we will primarily address the issues of biased reporting and the lack of statistical literacy in health professionals. Surprisingly, even doctors have trouble understanding medical evidence and are prone to being deceived by statistics, as we will demonstrate later.

### Biased Reporting in the Medical Literature

To explain what is meant by biased reporting in the medical literature, let us again take the example of the press release of the European trial for PSA screening. It stated that PSA testing reduces the risk of dying from prostate cancer by 20%. What does this number mean? It means that out of every 1,410 men who regularly participated in prostate cancer screening, one less died of prostate cancer than in an equally sized group of men who did not participate (Schröder et al. 2009). Additionally, 48 of the 1,410 men were unnecessarily treated and hence subjected to potential incontinence and impotence and that overall mortality was also unaffected. Communicating risk as a 20% risk reduction or as the number of men needed to screen to save one life makes quite a difference.

Evaluations of abstracts in leading medical journals have shown that the majority of reports fail to report absolute risks in addition to relative numbers (Schwartz et al. 2006; Sedrakyan and Shih 2007; Gigerenzer et al. 2010). Another form of biased risk communication in the medical literature is mismatched framing: Benefits are presented in relative risk reduction formats and appear rather large, whereas side effects are presented in absolute terms and appear smaller. Thereby benefits are overestimated, side effects underestimated. An even more extreme way to misinform is the omission of any side effects.

### Biased Reporting in the Media and Pamphlets

In contrast to presenting deceiving numbers, many health pamphlets do not present any numbers at all. A pamphlet informing the public about the human papillomavirus (HPV) vaccine – an innovative vaccine to prevent the risk of cervical cancer – states the following: "For two years,

**Table 24.1**

**The seven sins in health care:** The table summarizes the seven factors that contribute to misinformed patients identified by Gigerenzer and Gray (2011)

The seven sins	Example
<i>Biased funding of research</i>	Out of estimated US\$160 billion spent on research and development in health in the United States, more than half was sponsored by the pharmaceutical, biotechnical, and medical technology industry (see Gigerenzer and Gray 2011)
<i>Biased reporting in medical journals</i>	Out of 222 articles published in leading medical journals between 2003 and 2004, 150 failed to report the underlying absolute risk in the abstract (see Schwartz et al. 2006)
<i>Biased reporting in health pamphlets</i>	Out of 27 pamphlets informing about breast cancer screening in Germany, only ten informed about lifetime risk of developing breast cancer, two reported about risk reduction of death from breast cancer in relative formats, two in absolute formats, and one presented number needed to treat (see Kurzenhäuser 2003)
<i>Biased reporting in the media</i>	Out of 202 German Web sites and newspaper reports informing the public about HPV vaccination, 116 reported information about baseline risk of developing cervical cancer (96 gave correct estimates); 102 out of the 202 reports reported about pros and cons of the vaccination in a balanced way. Correct estimates of risk reduction were provided in 14 articles only (see Bodemer et al. submitted)
<i>Commercial conflicts of interest</i>	After a drug has been approved, doctors are offered money for each patient they put on the drug by companies (between 10 and 1,000€ per patient in Germany). In 2008, out of 150,000 private medical practices, about 85,000 participated in such programs (see Gigerenzer and Gray 2011)
<i>Defensive medicine</i>	In Switzerland, 41% of the general practitioners and 43% of internists reported that they sometimes or often recommend PSA tests for legal reasons. In other words: They order a test for patient which they would not order for themselves (see Steurer et al. 2009)
<i>Doctor's lack of understanding health statistics</i>	Ninety-six out of 160 gynecologists overestimated the positive predictive value of mammography and 29 underestimated the value, despite the fact that all relevant information was available (see Gigerenzer et al. 2007)

young women have had the possibility to get vaccinated against HPV. Worldwide, 50 million vaccines have been administered. In Germany, the media has reported controversy about the vaccine, while doctors and scientists are convinced of the certainty and efficacy of the vaccine.” What does this statement tell a young girl or her parents who are considering having her vaccinated? Does it mean that the vaccine reduces the risk of suffering from cervical cancer by 100%? Does the vaccine cause no side effects? Does the protection last a lifetime?

The case of the HPV vaccine is exemplary and of particular interest for two reasons: First, vaccination campaigns affect large parts of the population, primarily young girls between the ages of 12 and 16. Second, the HPV vaccine has prompted extensive media coverage, because some researchers have questioned whether it has been sufficiently evaluated (Dören et al. 2008). While the pamphlet conveys certainty and cites trustworthy and convinced experts, this

only reflects half the story. We conducted a media analysis in two countries – Germany and Spain – to evaluate media coverage and how the public was informed about the HPV vaccine in Web sites and newspapers. Most of the media reports did not provide any information about the prevalence, etiology, efficacy, or uncertainties of the vaccine (Bodemer et al. [submitted](#)). It is clear through content analyses of other health communications, such as pamphlets about mammography (Kurzenhäuser [2003](#)) or colon cancer screening (Steckelberg et al. [2001](#)) as well as media reports about medications (Moynihan et al. [2000](#)), that the media lack complete and balanced statistical information about risks, benefits, harms, and costs. Numbers are either not provided at all or are provided in nontransparent formats that mislead the public. This is alarming, since the mass media constitute the most prominent channels of communication about health innovations and treatments to the public (Grilli et al. [2009](#)).

## **Consequences of Biased Reporting**

Biased reporting undermines shared decision making and has consequences for the individual patient as well as for the health system. When the UK Committee on Safety for Medicine stated that the risk of life-threatening blood clots in legs or lungs is increased by 100% when using the third generation of the oral contraceptive pill, the public was appalled. As a consequence, many women stopped taking the pill, which resulted in unwanted pregnancies and abortions. But what did this 100% actually mean? Studies revealed that instead of 1 in 7,000 women who took the second generation of the contraceptive pill suffering blood clots, 2 in 7,000 who took the third generation pill did. This is equivalent to a relative increased risk of 100%, which in absolute numbers corresponds to a risk increase in 1 in 7,000 (example taken from Gigerenzer and Gray [2011](#)).

Another example is the fact that treatment benefits are often overestimated. When women and men in nine European countries were asked to estimate the effect of PSA screening and mammography on prostate cancer and breast cancer mortality reduction, respectively, they highly overestimated the benefits. Especially those who consulted their doctors or health pamphlets were particularly prone to overestimation (Gigerenzer et al. [2009](#)).

Another example is the consequence of false-positive test results, the fact that a test can erroneously signal a disease. False-positive tested patients often receive follow-up care despite the absence of disease, a phenomenon called overtreatment. Lafata et al. ([2004](#)) estimated incremental costs for false-positive results averaged over different screenings to be \$1,024 for men and \$1,171 for women, respectively, in the year following diagnosis. Moreover, besides unnecessary costs, false-positives create unwarranted anxieties and fears among patients.

These are just three examples that illustrate the dramatic consequences of biased reporting for health decisions. We will use these and other examples to better demonstrate and contrast the influence of different formats of risk communication.

## **How Good Are Experts and Laypeople at Dealing with Risks and Uncertainties?**

---

The public is often described as lacking the fundamental skills to deal with numerical information. Two terms have been coined to illustrate this phenomenon: (collective) “statistical illiteracy” (Gigerenzer et al. [2007](#)) and “innumeracy” (Paulos [1988](#)). Both concepts refer

to the widespread inability to understand quantitative information and to perform basic mathematical operations. But why is statistical literacy and numeracy so important for health decisions? Lipkus and Peters (2009) defined six main functions of numeracy that directly affect health decisions: Numeracy facilitates computation, encourages information search, improves interpretation of numerical information, facilitates the assessment of likelihood and value, can increase or decrease involvement in numerical data, and can consequently promote behavioral change.

## Statistical (Il-)literacy in Health

Gigerenzer et al. (2007) defined 13 principles of minimal statistical literacy. One of the key competences is the ability to deal with uncertainty. People tend to sustain an illusion of certainty – an ignorant perspective in a world that cannot guarantee any certainty at all (Gigerenzer 2002). For instance, when people rated which of five tests (DNA, fingerprint, HIV, mammography, expert horoscope) yield absolutely certain results, the majority (78%) believed that DNA tests do so. Furthermore, 63% believed in the certainty of fingerprint and HIV-test results, 44% stated that mammography leads to certain outcomes, and 4% even believed in horoscopes (Gigerenzer et al. 2007). One might think that experts are not prone to this illusion, but the opposite is true. In an undercover study, a client, who was explicit about not belonging to a risk group, asked 20 professional AIDS counselors the following questions in the mandatory pretest counseling session: Could I possibly test positive if I do not have the virus? And if so, how often does this happen? The vast majority stated that the test could not err and that it was absolutely impossible to receive false-positive results, which is, of course, not true, even though false-positives are rare (Gigerenzer et al. 1998).

Therefore, the first step to becoming statistically literate is to abandon this illusion and accept living with uncertainty. Minimal statistical literacy in health also subsumes an understanding of basic statistical concepts such as sensitivity, specificity, transforming conditional probabilities into natural frequencies, and the possibility of false alarms in medical screening tests as well as an understanding of the magnitude of treatment effects. All these concepts will be explained in this chapter. In addition, statistical literacy encompasses a grasp of the quality of scientific evidence and potential underlying conflicts of interests in medical research. For instance, the gold standard for evaluating a medical treatment is a randomized control trial (RCT). However, for many medical treatments no RCT is available and scientific evidence is inconclusive or even conflicting. Patients need to distinguish between different qualities of medical evidence. Another crucial distinction for decision making in health addresses the perspective from which a risk is evaluated. First, imagine a woman who knows that of 100,000 like her, 15 will have cervical cancer. She might decide not to participate in pap smear screening to identify early stages of cervical dysplasia since her baseline risk is rather low. Now, imagine a health policy decision maker: The pap smear screening reduces the annual incidence of cervical cancer in Germany by a total 10,400 women (Neumeyer-Gromen et al. *in press*). In this case, a national program to implement pap smear screening might be appreciated. Thus, depending on which perspective is taken, the evaluation of a treatment has different implications.

## The Concept of Numeracy

---

The second approach to assessing people's ability to deal with mathematical concepts is numeracy. In a broader sense, numeracy is defined as "the aggregate of skills, knowledge, beliefs, dispositions, and habits of mind – as well as the general communicative and problem-solving skills – that people need in order to effectively handle real-world situations or interpretative tasks with embedded mathematical or quantifiable elements" (Gal 1995, cited in Reyna et al. 2009). A more concrete definition of health numeracy is given by Golbeck et al. (2005): "the degree to which individuals have the capacity to access, process, interpret, communicate, and act on numerical, quantitative, graphical, biostatistical, and probabilistic health information needed to make effective health decisions." Some also subsume the ability to read and understand graphs under the term "health numeracy" (e.g., Ancker and Kaufman 2007), but we use the term "graph literacy" to define the ability to use visualizations (Galesic and Garcia-Retamero 2010). Moreover, Golbeck et al. (2005) differentiate four levels of health numeracy: Basic health numeracy encompasses the ability to identify numbers and correctly interpret quantifications. Computational health numeracy includes the ability to count and to conduct simple manipulations of numbers and quantities. Concepts of inference, estimation, proportions, frequencies, and percentages are represented on an analytical level of health numeracy. Finally, statistical health numeracy involves an understanding of biostatistics, the ability to compare numbers on different scales, and the critical analysis of risk ratios or life expectancy. Similar to statistical literacy, health numeracy also incorporates the understanding of scientific concepts, such as randomization and the double-blind study. Likewise, Reyna et al. (2009) reviewed the literature on numeracy and defined three levels of numeracy: The lowest level covers concepts of the real number line, time, measurement, and estimation. The middle level requires simple arithmetic operations and the comparison of magnitudes, while the highest level consists of an understanding of ratios, fractions, proportions, percentages, and possibilities.

## Measuring Numeracy

Different measures have been developed to assess people's numeracy skills. Objective scales assess competence with items that measure basic, computational, analytical, or statistical abilities. For example, a simple three-item scale by Schwartz et al. (1997) requires the conversion of percent into proportion and vice versa and the estimation of the expected numbers of heads in 1,000 coin tosses. This scale was the basis for an 11-item numeracy scale developed by Lipkus et al. (2001). An alternative way to measure numeracy is with the Subjective Numeracy Scale, which asks subjects to indicate their confidence in their own mathematical skills and preferences for numerical versus verbal risk information (Fagerlin et al. 2007). Subjects have to rate how easily they can calculate a 15% discount on a T-shirt or whether they prefer weather forecasts that state a probability of rain (e.g., 20% chance of rain tomorrow) as opposed to a verbal description (e.g., a small chance of rain tomorrow). The advantage of the Subjective Numeracy Scale is that subjects are not tested but rather are allowed to estimate their own abilities and preferences. The scale showed satisfactory correlations with objective scales and is easy to apply (Zikmund-Fisher et al. 2007; Galesic and Garcia-Retamero 2010).

## Numeracy in Experts and Laypeople

So how widespread is innumeracy? The Programme for International Student Assessment (PISA) in 2003 assessed mathematical and problem-solving skills of 15-year-olds in 24 countries. The results revealed low mathematical literacy skills in the United States and Germany – especially in concepts such as uncertainty and quantity. In 2007, the National Assessment of Educational Progress (NEAP) assessed students' mathematical performance. Only 22% of the students at grade 12 performed at a proficient level or above; 37% performed at basic level, and 41% even below basic level (Grigg et al. 2007; for an overview see Reyna et al. 2009). However, these results are not surprising since statistics and probability calculation are rarely implemented in school curricula. Nor is it surprising that adults have similar major difficulties in performing simple computations. The National Adult Literacy Survey (NALS) includes one scale measuring quantitative abilities. It demonstrated that 47% of the adults surveyed had very low quantitative literacy scores and difficulties in performing simple mathematical operations (Kirsch et al. 2007). These results were replicated by the National Assessment of Adult Literacy in which 36% of the subjects had a maximum of basic quantitative abilities (Kutner et al. 2006). Galesic and Garcia-Retamero (2010) compared numeracy skills in Germany and the United States using national probabilistic samples. On average, numeracy skills were higher in Germany (average proportion of correct items: 68.5% vs. 64.5%) with a greater difference between literate and illiterate in the United States. In other studies, even in well-educated samples only 16–25% of the subjects gave correct answers to all three items of the short numeracy scale (Lipkus et al. 2001; Schwartz et al. 1997). In general, men achieve higher scores than women, younger people higher scores than older people, and more educated people higher scores than those less educated.

## Consequences of Innumeracy

A growing body of literature has revealed consequences of the lack of statistical literacy and numeracy skills. In one study, women read data about mammography screening and breast cancer mortality and assessed their personal risk of dying of breast cancer with and without screening. Women with low numeracy skills (none of the three items in the short numeracy scale answered correctly) had an accuracy rate of 5.8%; in comparison, those women with high numeracy skills (3 of 3 items correct) showed an accuracy rate of at least 40% (Schwartz et al. 1997). In another study, subjects were confronted with the baseline risk of a hypothetical disease and had to choose between two treatments. Benefits of the treatments were presented as number needed to treat, relative risk reduction, absolute risk reduction, or a combination of these formats. Independent of the format, high-numeracy subjects were more successful in identifying the more beneficial treatment and correctly calculating the effect of treatment for a given baseline risk than less numerate subjects (Sheridan et al. 2003). Low-numeracy subjects were also more prone to framing effects (Peters et al. 2006). Treatment effects can either be framed positively by stating that 80 of 100 patients survive a treatment or negatively by stating that 20 of 100 patients actually die. Differences between the two frames affect decisions, more so in less numerate subjects than in highly numerate students. In addition, less numerate people have more difficulties transforming one representation format (e.g., frequency “20 of 100”) into another (e.g., probability “20%”): Whereas highly numerate people give consistent

risk estimations independent of the format, less numerate people give lower risk estimates under probability than frequency formats. People low in numeracy also tend to overestimate their personal risks, which in turn has important consequences for the perception of treatment benefits and treatment decisions (Woloshin et al. 1999; Davids et al. 2004; Dieckmann et al. 2009). Finally, numeracy moderates denominator effects. People low in numeracy tend to ignore the information in the denominator, which leads to the misinterpretation of treatment effects when the sample size in the treatment and control group are unequal (Garcia-Retamero and Galesic 2009).

On a behavioral level, patients show difficulties in disease management. For instance, diabetes patients low in health literacy – the ability to perform the basic reading tasks needed to function in the health-care environment – and numeracy showed a poorer anticoagulation control (Estrada et al. 2004). Rothman et al. (2006) investigated the perception and interpretation of food labels in 200 primary care patients. Even though most patients indicated that they frequently used food labels and stated that these labels are generally easy to understand, many patients misunderstood information about serving size, misapplied extraneous material on the food label, and performed incorrect calculations.

## The Role of Numbers and Words in Risk Communication

---

The communication of risks does not necessarily require an understanding of numerical information. Instead of relying on statistics, information about treatment benefits or harms can be based on the experiences of doctors and patients. Furthermore, verbal probability estimates describe risks without using data. In the following, we will describe the discrepancy between statistical and narrative evidence, and the influence of verbal probability estimates as opposed to numerical probability estimates on risk perception.

### Narrative Versus Statistical Evidence

---

Imagine a woman age 53 must decide whether to participate in mammography screening. She decides to ask her doctor about the test. The doctor gives her the following information: "Here is what we know: Think about two groups of women at age 40 or older. In each group are 2,000 women. Whereas one group receives biannual mammography screening, the other group does not receive any screening. After 10 years, the breast cancer mortality in the two groups is compared. In the screening group, 5 out of 2,000 women died of breast cancer, whereas in the control group, 6 out of 2,000 died of breast cancer. Mammography screening prevented 1 breast cancer death out of 2,000 women." The woman is not convinced to participate in the screening. On her way back home, she meets her neighbor – a 62-year-old woman. She asks her whether she has ever participated in mammography screening and receives the following answer: "Oh, yes, fortunately, I did. About 6 years ago, my doctor advised me to have a mammogram. At that time, I didn't really know what it was and didn't know a lot about breast cancer either. But I thought it couldn't harm and did it. Then, the mammogram turned out to be positive. Of course, I was shocked. But the doctor told me that my chances are very good, since the cancer was detected at an early stage. I had a mastectomy, and since then, I'm doing fine. You can imagine how happy I am that I had the mammogram." After talking to her neighbor, the woman is convinced – she will make an appointment for a mammogram tomorrow.

This example illustrates two different types of evidence a decision maker is often confronted with: statistical evidence and narrative (anecdotal) evidence. Whereas the latter usually encompasses stories and experiences from single cases ( $N=1$ ), the former summarizes data of larger samples ( $N > 1$ ). Which of the two types of evidence is more persuasive? Reinard (1988) reviewed the literature on statistical and narrative evidence and found little support for an advantage of statistical messages over narrative messages. Anecdotes and stories are more vivid, lively, and emotionally charged (Nisbett and Ross 1980; Taylor and Thompson 1982) or in other words, “it is generally accepted that stories are more concrete, more imagery provoking, and more colorful than statistics that are often abstract, dry, and pallid” (Baesler and Burgoon 1994). Consequently, narrative evidence increases personal relevance, especially when the receiver can identify with the narrator – as the 53-year-old woman did with her neighbor. In contrast, the statistical evidence provided by the doctor appears abstract and less imagery provoking. However, statistical evidence offers some advantages over anecdotes that are of particular importance for a decision maker. Statistical evidence provides information about baseline risks and treatment benefits based on a larger sample size. In comparison to stories, statistics are more factual, objective, and scientific and thereby establish a basis for credibility and trust (Baesler 1997). In their meta-analysis, Allen and Preiss (1997) also found a slightly more persuasive effect of statistical evidence than narrative evidence in different settings.

Both statistics and narratives are common in medical decision making. Ubel et al. (2001) asked subjects to choose between two treatments for angina – bypass surgery and balloon angioplasty. Both kinds of evidence were presented to the decision maker: Statistical evidence for bypass surgery showed a 75% success rate and balloon angioplasty a 50% success rate. When the narrative evidence was proportionate – in other words, when it reflected the statistical success rates (i.e., three pro statements and one contra statement for bypass surgery, and one pro and one contra statement for balloon angioplasty), 44% selected bypass surgery. However, when the number of narratives was disproportionate (one pro and one contra statement in both conditions, independent of the success rate), only 33% favored bypass surgery. Even though both conditions included statistical information, the proportion of narrative evidence affected treatment choice. In a second study, four testimonials were always presented, either proportionate or disproportionate. Whereas no significant difference in treatment choice was found between the proportionate and disproportionate format (37% and 34% chose bypass surgery), many more (58%) chose bypass surgery when numerical information was given without narratives.

The use of statistical or narrative evidence also influences risk perception. Subjects receiving a narrative reported higher personal risk than those who received statistical information, as well as higher intentions to get vaccinated, when confronted with a decision about vaccination against the hepatitis B virus (deWit et al. 2008).

In sum, the issue of which kind of evidence is more persuasive is still unresolved. As we pointed out at the beginning of this chapter, persuasion might not be an appropriate goal when conveying health messages. Instead, correct risk estimates as well as trust and credibility reflect more central evaluation measures. Since different evidence formats affect health decisions, it is crucial to understand how people perceive and interpret narrative and statistical information. Let us again consider the example of the woman facing the mammography screening decision. If she ignores the statistical evidence, she might erroneously assume that mammography is certain and prevents breast cancer deaths by 100%. Statistics help to objectively convey treatment benefits and harms and thereby help to inform and educate patients.

## Expressing Probabilities with Words Versus Numbers

Risk information can, but does not necessarily have to include numbers. A meteorologist may predict a “10% chance of rain” or alternatively state that it is “unlikely” to rain tomorrow. Similarly, a physician can tell a patient that it is “very probable” that she will recover from the treatment, instead of stating that her “chances are 80%” (or in other words that 8 out of 10 patients recover). Both formats represent options for risk communication – but which is more transparent and informative? Words are more common in communication than numbers and therefore match people’s internal representation, whereas the concept of probability emerged rather late in human history (Hacking 1975; Zimmer 1983). In addition, verbal probability expressions signal vagueness and uncertainty since words can never be as precise as numerical point estimates. At the same time, the imprecision of a verbal probability is its main flaw: People show immense variation in the interpretation of verbal probabilities (Budescu and Wallsten 1985; Brun and Teigen 1988). Brun and Teigen (1988) investigated how people interpret verbal probabilities and found high between-subject and within-subject variability in different domains. For instance, subjects assigned lower numerical estimates to verbal probabilities in a medical context in contrast to a context-free condition (see also Pepper and Prytulak 1974). One potential explanation of context dependency in the interpretation of verbal probability estimates is perceived base rate (Wallsten et al. 1986). A higher numerical probability estimate was associated with a verbal probability expression when the base rate of the event was high. This effect occurred primarily in verbal expressions of high and medium probability terms (e.g., possible, very likely), less so in low probability terms (e.g., rarely). Likewise, Weber and Hilton (1990) discussed perceived personal base rate and perceived severity as factors that influence the interpretation of verbal probabilities. Probability ratings were higher for more severe events, even when controlled for the base rate effect.

### Verbal and Numerical Probabilities in Health

Verbal probability estimates are common in health, especially in doctor–patient communications. Doctors often describe risks and treatment effects with such verbal expressions as unlikely, probable, or certainly. However, what a doctor means by “probable” is not necessarily what a patient understands by the same term. When ranking eight different probability expressions, mothers showed higher interquartile ranges than doctors, meaning that the range of interpretation of a single expression was larger for laypeople than for experts (Shaw and Dear 1990). Can the implementation of standards in medical risk communication reduce this discrepancy? The European Commission (1998) established guidelines to indicate the frequency of side effects with five verbal terms, each representing a particular frequency (see  Table 24.2). Knapp et al. (2004) compared how laypeople estimated the side effects in a verbal estimate condition and a numerical estimate condition with two different side effects of statins. One side effect, constipation, had a risk of 2.5%, which corresponds to a “common” event according to the guidelines; the other side effect, pancreatitis, had a risk of 0.4%, which is described as “rare.” Subjects had to rate the likelihood that they would experience the side effect. The average estimated probability of occurrence for the common side effect constipation was 34.2% in the verbal condition and 8.1% in the numerical condition. For pancreatitis, the estimates were 18% in the verbal and 2.1% in the numerical condition. In general, patients give

**Table 24.2**

**Verbal versus numerical probability estimates:** Verbal probability estimates and their intended numerical equivalent from the European guideline on the readability of the label and package leaflet of medical products for human use (1998). When verbal probability estimates are presented without numerical information, laypeople tend to overestimate the occurrence of side effects. In other words, the verbal descriptors are interpreted differently by laypeople than intended by the guidelines (see Steckelberg et al. 2005)

Verbal probability estimate (proposed by European guidelines)	Numerical probability estimate (intended by European guidelines)	Estimated probability by laypeople (Mean [SD])
Very common	>10%	65 (24)%
Common	1–10%	45(22)%
Uncommon	0.1–1%	18(13)%
Rare	0.01–0.1%	8(8)%
Very rare	<0.01%	4(7)%

higher estimates for verbal probabilities than actually intended by the guidelines, which in turn influences risk perception and behavior (Berry et al. 2004).

Marteau et al. (2000) tested parents in their understanding of test results for prenatal diagnostics. When a numerical format for the test outcome was used, 97% interpreted the result correctly, whereas only 91% did so when verbal probabilities were given. Gurmankin et al. (2004) compared variations in risk perception when subjects in a hypothetical cancer scenario received either a verbal message only or a verbal message plus numerical information. In general, subjects overestimated their relative risk and showed very high variation in their estimates, within and between subjects.

### Preferences for Verbal Versus Numerical Probabilities

Independent of how people interpret and understand verbal or numerical probability estimates, they might have a preference for one of the two formats. Mazur et al. (1999) confronted male patients with the treatment choice of either watchful waiting or surgery now in a prostate cancer scenario. The treatment effect in the surgery-now option was described as “possible,” whereas side effects were presented in numerical information (i.e., 10–25% chance of total loss of bladder control after surgery). More than half of the patients (56%) preferred numerical information. Those patients who preferred numbers chose watchful waiting more often, compared to those preferring verbal risk information. Similarly, Shaw and Dear (1990) asked parents which format doctors should use to communicate risks and found that 72% felt that they understood the numerical information and 66% actually favored doctors who gave numerical estimates. In general, findings suggest that people tend to prefer probability information in numerical formats when they search for information but use verbal probabilities when they communicate risks to others (see, e.g., Erev and Cohen 1990; Wallsten et al. 1993). Potential reasons for the preference for numerical information is that people trust numerical

information more and feel more comfortable and satisfied than with verbal estimates (Berry et al. 2004; Gurkankin et al. 2004). Despite this general pattern, interindividual differences exist. Some people feel uncomfortable with numbers and shrink from statistics, while others actively search for numerical information.

In sum, statistical and narrative evidence are important sources for decision makers but at the same time affect risk perceptions and decisions differently. People tend to perceive numbers as objective and credible. Verbal estimates lead to high inter- and intraindividual variation in the interpretation of risks. However, the “strength” of statistical evidence depends on two crucial factors. First, we demonstrated that people often lack statistical literacy and numeracy skills. Even if the public prefers to base decisions on statistics, can it adequately understand them? Low numeracy results in misconceptions that undermine informed decisions. The elimination of statistical illiteracy and innumeracy requires educational programs for doctors, patients, and children to establish a risk-literate society.

The second factor refers to a very different problem. The problem of risk communication is not simply in people’s minds – their inability to deal with numbers – but rather in the environment – an environment that is primarily characterized by biased and nontransparent communication formats. Different representation formats exist to express the same (numerical) information, for example, frequencies (20 of 100) and percentages (20%). Some formats mislead people and lead to false expectations. Other formats are rather intuitive and make it easy for recipients to correctly assess a risk. What makes a format transparent is its ecological structure: the match of the external representation format and the mind – that is, the cognitive capacities to recognize relationships in certain representations of complex problems (Gaissmaier et al. 2007). The second part of this chapter will focus on this issue: transparent risk communication formats and how they facilitate the interpretation of numbers.

## **Transparent Risk Communication: How to Overcome Statistical Illiteracy and Innumeracy**

---

The problem of biased risk communication is less in people’s lack of statistical competency, but primarily in the use of nontransparent communication formats. We will present shortcomings of relative risks, conditional probabilities, 5-year survival rates, and their transparent counterparts. Additionally, we will illustrate the potential benefits of graphical representations and discuss approaches to including uncertainty in risk communication.

### **Relative Versus Absolute Risks**

---

Let us refer again to the example of the UK pill scare. When the UK Committee on Safety for Medicine stated that the risk of life-threatening blood clots in lungs and legs increased by 100%, many women stopped taking the pill. The consequences were unwarranted pregnancies and abortions. Although stating a 100% increase is not incorrect, if an absolute instead of a relative format of risk increase is used (the risk increased by 1 in 7,000 women – i.e., instead of one woman, two women in 7,000 had blood clots), the risk appears to be very different.

A relative risk is the ratio of a risk in a treatment group and the risk in a control group who did not receive a treatment (or received only a placebo). The relative risk reduction is simply

calculated by subtracting the relative risk from one. An absolute risk is defined by the difference in absolute magnitudes between the two groups. A third format to express the same information is the number needed to treat, that is, the number of patients who have to be treated to prevent one death (e.g., 100 people have to get vaccinated to prevent one death). In principle, the three measures can be converted into each other if the underlying risk is known. One might think these formats can be interchangeably used in risk communication – but the opposite is true. As the pill scare example demonstrates, the perception of a treatment's risk increase highly depends on the presented format.

Malenka et al. (1993) asked patients to select one of two treatments for a hypothetical disease with equivalent efficacy, side effects, and costs. The only difference was that one medication was framed in terms of relative risk reduction and the other as absolute risk reduction. The majority of patients (56.8%) selected the medication with relative numbers; about 15% were indifferent, and about the same proportion selected the medication with absolute numbers; and 13% could not decide. Similarly, Sarfati et al. (1998) showed subjects three different (fictitious) screening programs, each in a different format – relative risk reduction, absolute risk reduction, or number needed to treat. Depending on the format, subjects' willingness to participate differed substantially. When framed as relative risk reduction, 80% intended to participate, in comparison to only 53% and 43% who did so in the absolute risk reduction and number needed to treat condition, respectively. Additionally, relative risk reduction formats lead to higher deviations in treatment decisions from expected-utility theory assumptions than absolute formats (Hembroff et al. 2004). Does the same hold true when the subjects are medical experts? Naylor et al. (1992) showed that information in the form of relative risk reduction (relative decrease of 34%) led to higher perception of treatment effects in doctors compared with absolute risk reduction (decrease from 3.9% to 2.5%) or number needed to treat (77 persons have to be treated to save one patient). Likewise, doctors' mean ratings for effectiveness of a drug that lowers cholesterol concentration depended on whether relative or absolute risk reductions were presented (Bucher et al. 1994). A diabetes prevention intervention was rated as important or very important by 86% of health professionals under a relative risk format condition, whereas only 39% gave these ratings in an absolute risk reduction condition (Mühlhauser et al. 2006).

Consistent with other reviews (e.g., Edwards et al. 2001; Moxey et al. 2003) a meta-analysis by Covey (2007) supported the conclusion that both laypeople and experts are sensitive to the way risk reduction is framed: People perceive higher treatment effects when relative risk reduction formats are used in comparison to absolute risk reduction and number needed to treat.

As previously mentioned, health reporting is most biased when different formats are used for different effects, known as mismatched framing. Describing treatment benefits in relative numbers and treatment harms in absolute numbers confuses and misleads patients. For instance, a German pamphlet about hormone replacement therapy (HRT) states the following: 60 out of 1,000 women develop breast cancer in their lives. After HRT for 10 years, 66 out of 1,000 women develop breast cancer – the absolute increase is 6 in 1,000. At the same time, only half as many of the women who take HRT develop colon cancer, compared to those who do not take HRT; in other words, HRT reduces the risk of developing colon cancer by 50%. By using two different formats to describe benefits and harms of HRT, the consumer is misled and overestimates the benefits in contrast to the harms.

However, some argue in favor of the use of relative risks and odds ratio, especially in meta-analyses. The rationale is that both formats are supposedly more stable across different

subpopulations than absolute risks (e.g., Smeeth et al. 1999). In any case, this does not have any implications for risk communication, which should always be based on absolute numbers.

Lessons learned: Findings demonstrate that no relative risks should be used in risk communication. Risk reduction or risk increase ought to be presented in absolute numbers only.

## Conditional Probabilities Versus Natural Frequencies

We already mentioned the illusion of certainty and the problem that patients and health professionals often believe in the certainty of medical test results. For instance, 44% in one study stated that the result of a mammogram is absolutely certain (Gigerenzer et al. 2007). But what is actually the probability that a woman has breast cancer given a positive mammogram? To illustrate what a positive mammogram means, look at the following information (Eddy 1982; see Gigerenzer et al. 2007):

- The probability of breast cancer is 1% for a woman at age 40 who participates in routine screening (this is the prevalence or base rate).
- If a woman has breast cancer, the probability is 90% that she will have a positive mammogram (this is the sensitivity or hit rate).
- If a woman does not have breast cancer, the probability is 9% that she will also have a positive mammogram (this is the false-positive rate).

The task is to estimate the probability that a woman at age 40 who had a positive mammogram actually has breast cancer. What is the correct answer? When Eddy (1982) presented a similar scenario to staff at the Harvard Medical School, 95 of 100 physicians gave an answer between 70% and 80%, though the correct answer is about 10% – or in other words, of ten women who had a positive mammogram, only about one actually has breast cancer. Why do people have problems solving this and similar tasks?

The simple answer would be that humans are not “Bayesian” and hence are not capable of calculating posterior probabilities  $P(H|D)$  based on the prior probability  $P(H)$ , the likelihood  $P(D|H)$ , and the probability  $P(D|-H)$ . In the mammography example,  $P(D|H)$  is the sensitivity (90%),  $P(H)$  is the base rate (1%), and  $P(D|-H)$  is the false-positive rate (9%). The computation of the posterior probabilities requires Bayes’s theorem:

$$P(H|D) = \frac{P(D|H) \cdot P(H)}{P(D|H) \cdot P(H) + P(D|-H) \cdot P(-H)} \quad (1)$$

Kahneman and Tversky (1972) stated that humans cannot perform Bayesian reasoning and lapse into cognitive biases. For instance, humans tend to ignore base rates when calculating conditional probabilities. As a consequence of these biases, people’s judgments are often inconsistent with normative “Bayesian” prescriptions (Casscells et al. 1978; Eddy 1982). Erroneously, humans do not differentiate between the probability of a disease given a positive test result (the posterior probability), and the probability of having a positive test result given the disease (sensitivity). Bayes’s theorem is the common formula in most medical and statistical textbooks to calculate posterior probabilities, but still people who should be familiar with the formula seem to have difficulties in its application.

Gigerenzer and Hoffrage (1995) challenged the assumption that people cannot solve Bayesian tasks and proposed an alternative representation format that facilitates the

computational process: natural frequencies. Think again about the mammography example, but this time, the following information is given:

- Ten of 1,000 women at age 40 who participate in mammography screening have breast cancer (prevalence or base rate).
- Of these ten women, nine have a positive mammogram (sensitivity or hit rate).
- Of the 990 women who do not have breast cancer, about 89 will have a positive mammogram nonetheless (false-positive rate).

Now imagine a representative sample of 1,000 women aged 40 who participate in breast cancer screening. How many of these women with a positive test result actually have breast cancer? Of course, the answer is the same: About 1 of 10. Nevertheless, to arrive at the correct solution does not require Bayes's theorem. Instead, the calculation is much simpler: Of 1,000 women, 98 will have a positive mammogram (9 of the 10 women who actually have breast cancer – referred as  $a$  in the formula and 89 of the 990 healthy women, referred as  $b$  in the formula). Of this 98 with a positive test result, only 9 actually have breast cancer, which results in 9.2%, or about 10%.

$$P(H|D) = \frac{a}{a+b} \quad (2)$$

When Gigerenzer and Hoffrage presented the mammography problem in natural frequencies instead of conditional probabilities, about half of the subjects gave the correct solution in comparison to only one-quarter in the probability condition. Since then, many researchers have replicated the results. Cosmides and Tooby (1996) conducted a series of experiments and supported the hypothesis that natural frequencies lead to higher proportions of correct inferences compared to probability formats. For example, they replicated Casscells et al. (1978) study: Only 12% of their subjects arrived at the correct result when confronted with probabilities, but between 56% and 76% did so when confronted with natural frequencies.

Nevertheless, the concept of natural frequencies has aroused controversy about whether and why it facilitates Bayesian reasoning. Some researchers have argued that frequencies per se do not improve people's performances in Bayesian tasks. However, they confused natural frequencies with other kinds of frequencies (for an overview see Hoffrage et al. 2002). For instance, Macchi and Mosconi (1998) demonstrated that not all kinds of frequencies facilitated Bayesian reasoning, and Lewis and Keren (1999) reached a similar conclusion. Gigerenzer and Hoffrage (1995) stated in the original paper that the computational simplification can be obtained only for natural frequencies, not for normalized frequencies, which – just like probabilities – require Bayes's theorem. Further misunderstandings have resulted from proposed alternative explanations for the same phenomenon, for example, that the facilitating effect is based on a “nested-set structure” or on “partitive representations” (Barbey and Sloman 2007), which actually just restate the original argument.

Barton et al. (2007) proposed a statistical taxonomy subsuming three orthogonal dimensions to reduce confusion: First, the information can refer to one event only (single-event probabilities) or a set of events (frequencies). Second, different numerical representations, such as percentages, fractions, real numbers between 0 and 1, and pairs of integers, are differentiated. Third, the information can be presented in normalized formats or nonnormalized (also called conjunctive) formats. Due to the orthogonality of the dimensions, any combination is possible. For instance, expressing the mammography information in chances leads to the same

computational effect as doing so with natural frequencies but refers to a single individual (Brase 2009).

In summary, findings support the conclusion that natural frequencies help people solve Bayesian tasks and understand positive predictive values. Doctors and patients can easily learn what a positive test result means and how prevalence, sensitivity, and false-positives interact. Teaching natural frequencies is also rather simple. Instead of Bayes's rule, which invites learners to forget the actual components of the formula, the principle of natural frequencies is easy to grasp and helps people convert probabilities into natural frequencies. Even children benefit from this representation format and can perform Bayesian tasks (Zhu and Gigerenzer 2006).

We already mentioned the consequences of not understanding positive predictive values in the  [Introduction](#). Imagine a woman aged 56 who has a positive mammogram. Besides being extremely worried and anxious after receiving this result, she has to undergo further examinations and treatments. However, the chances of her result just being a false-positive are 9 out of 10. The terms "overdiagnosis" and "overtreatment" have been coined to call attention to the phenomenon that many people who have a positive screening test result are actually treated, despite the absence of the disease.

Lesson learned: While people have difficulties interpreting and calculating conditional probabilities, natural frequencies facilitate Bayesian reasoning.

## Single-Event Probabilities Versus Frequencies

---

Another advantage of frequency statements is that they always include a reference class. This is not the case for single-event probabilities. A single-event probability is defined as "a probability that refers to an individual event or person" (Gigerenzer et al. 2007). Thus, no reference classes are included, which often leads to misconceptions between a communicator and a receiver. A meteorologist forecasts that the probability of rain tomorrow is 30%. This prediction leaves room for different interpretations. It could mean that it will rain 30% of the time, in 30% of the area, or on 30% of the days like the one tomorrow (Gigerenzer et al. 2005). While the latter interpretation is correct, most people believe the other two options to be true. Similarly, stating that the probability of developing sexual problems as a consequence of a drug is 30% leaves the patient alone to his or her subjective interpretation. Again, the probability could refer to 30 out of 100 sexual encounters of a single person or to 30 out of 100 patients taking the drug. Frequency statements always include a reference class and thereby eliminate misunderstandings.

Lesson learned: Always provide the reference class to which a probability refers.

## Five-Year Survival Rates Versus Mortality Rates

---

When evaluating a health treatment, the first question that comes to mind is whether it saves lives in the long run. Cancer screenings aim at identifying a cancer at an early stage, even before first symptoms occur. Thus cancer screenings usually increase incidence rates – the number of cancers in a given population within a given time frame. This fact prevents us from drawing conclusions about a screening's effects on life expectancy.

Probably the most common unit mentioned when evaluating health treatments is the so-called 5-year survival rate. Survival rates can be defined as the number of patients alive at a specified time following diagnosis (such as after 5 years) divided by the number of patients diagnosed (Gigerenzer et al. 2007).

5-year survival =

$$\frac{\text{Number of persons diagnosed with a specific cancer still alive 5 years after diagnosis}}{\text{Number of persons diagnosed with a specific cancer in the study population}} \quad (3)$$

For example, if a screening detects 100 people who have a positive diagnosis, and 80 of them are still alive after 5 years, the 5-year survival rate is 80%. If only 20 are still alive, the 5-year survival rate is 20%. One might expect that the higher the 5-year survival rate, the better. However, there is an alternative to 5-year survival rates: (annual) mortality rates. And even more surprising, the correlation between changes in 5-year survival rates and changes in mortality rates over time is zero (Welch et al. 2000). The mortality rate is defined as the number of people in a group who die annually from a disease, divided by the total number of people in the group.

$$\text{annual mortality rate} = \frac{\text{number of persons who die from a specific cancer over 1 year}}{\text{number of persons in the study population}} \quad (4)$$

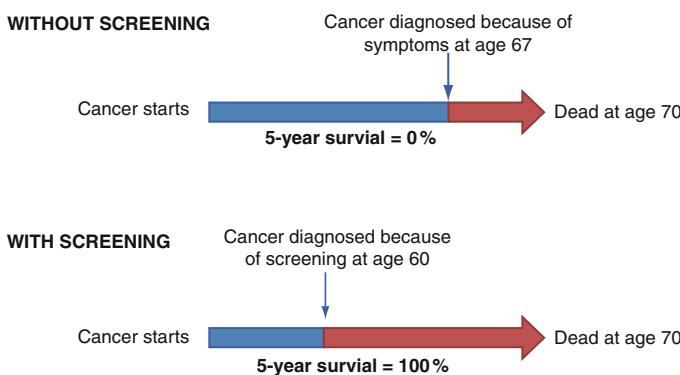
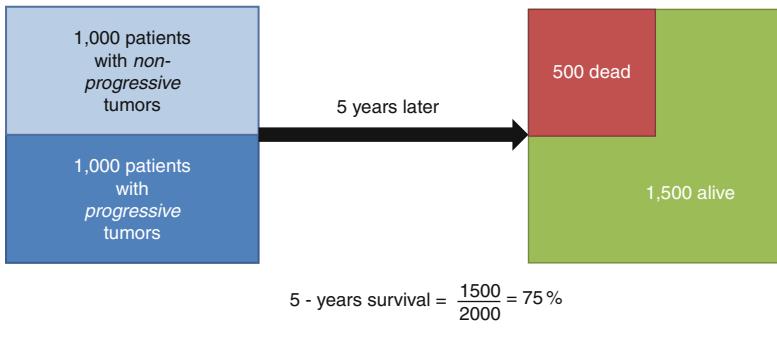
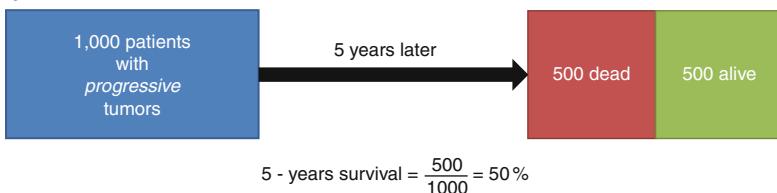
By comparing the two formulas, one thing is apparent: The 5-year survival rate only includes people who are *diagnosed* with disease in the denominator. This fact makes it prone to biases. An annual mortality rate, instead, includes the entire population at risk in the denominator. However, 5-year survival rates are the rule rather than the exception in current risk communication.

## Lead-Time Bias

The first shortcoming of 5-year survival rates is the so-called lead-time bias: A higher proportion of people alive 5 years after screening does not necessarily mean that people actually live longer – it might only be an illusionary extension of life (see Fig. 24.2a). Assume that the 5-year survival rate for a specific cancer was 1% between 1960 and 1965. For the time between 2005 and 2010, the proportion of people alive after 5 years was 80%. The cancer in the 1960s was diagnosed when patients first showed symptoms. In the years 2005–2010, a special screening test was applied for earlier detection, even before first symptoms occurred. Thus the time to diagnosis was reduced. However, this does not necessarily mean that it prolongs life, since patients between 1960 and 1965 and 2005 and 2010, respectively, might die the same number of years after developing the cancer. The difference is that those who lived between 2005 and 2010 knew earlier; in other words, they lived longer with the diagnosis, but in fact their life expectancy was the same.

## Overdiagnosis Bias

The second shortcoming is the so-called overdiagnosis bias. Not every tumor or dysplasia (that is detected) is necessarily fatal. Some tumors would have never been detected without screening, because they would never have caused any symptoms, would have resolved spontaneously,

**a****b****Fig. 24.2**

**Shortcomings of 5-year survival rates:** The figure illustrates the two potential biases of 5-year survival rates (modified from Gigerenzer et al. 2007). (a) *Lead-time bias:* The arrows illustrate the course from the beginning of a disease to death. In the group without screening, cancer is diagnosed at age 67, in the screening group at age 60. However, in both groups, patients die at the same age (age 70). Whereas in the non-screening group the 5-year survival rate is 0%, it is 100% in the screening group. (b) *Overdiagnosis bias:* (1) A group of 1,000 patients with progressive tumors is monitored over 5 years. After 5 years, 500 are still alive; the survival rate is 50%. (2) The same group of 1,000 patients with progressive tumors is monitored over 5 years. Additionally, the screening detects patients with nonprogressive, indolent tumors. Again, after 5 years 500 patients died (500 out of 1,000 with progressive cancer). However, in the calculation of the 5-year survival rate, those 1,000 with nonprogressive tumors are also included hence the 5-year survival rate is 75%

or would not have been detected before the patient died of other causes. To illustrate this example, take a look at the formula 3 for the 5-year survival rates. Assume that we have 1,000 people in a screening with a positive diagnosis of having a progressive tumor based on symptoms. These 1,000 patients form the denominator to calculate the survival rate. After 5 years, half of them are still alive. Hence, the survival rate is 50%. Now imagine that our screening also detects very small and indolent tumors. These tumors are nonprogressive and hence not lethal. To the 1,000 patients with progressive tumors in the denominator, add the 1,000 patients with indolent tumors. However, the number of deaths due to the cancer is still 500. Now the survival rate is 75%. Although the number of deaths remains the same, the 5-year survival rate provides a much more favorable picture – a bias that does not affect annual mortality rates (see [Fig. 24.2b](#)). One well-known example is the Mayo Lung Project of the 1970s and 1980s. Smokers were assigned to either a screening group or a control group receiving no screening. Whereas 206 lung tumors were detected in the screening group, only 160 were detected in the control group. However, the overall mortality in both groups was the same. A follow-up in 1999 with patients of both groups who were still alive and had no positive diagnosis in 1983 showed that 585 patients of the screening group compared to 500 in the control group had lung cancer. One interpretation of the data is that the screening detected small and indolent tumors that were not lethal and therefore did not need any treatment (Marcus et al. [2006](#)).

In summary, an increase in 5-year survival rates can be observed under three conditions: First, due to lead-time bias, the tumor is detected earlier, but patients still die at the same age as without screening. Second, due to overdiagnosis, more tumors are detected, but among them are indolent ones – the number of deaths is not affected. Third, screening allows for earlier detection and better treatment – in this case the screening is indeed beneficial. Mortality rates also capture the latter effect but do not fall into the traps of lead-time and overdiagnosis bias. As mentioned above, the correlation between changes in 5-year survival rates and changes in mortality rates is zero. Furthermore, a comparison of survival and mortality rates for 20 tumors between 1950–1954 and 1989–1995 showed an absolute increase in 5-year survival rates of between 3% and 50% over 5 years, whereas changes in mortality rates ranged from –80% to 259% (Welch et al. [2000](#)). As mentioned above, those changes were completely uncorrelated.

The influence on treatment evaluations of 5-year survival rates in comparison to mortality rates was shown by Wegwarth et al. ([2011](#)). They presented physicians with either a 5-year survival rate only, annual disease-specific mortality only, or a combination of the two formats with or without incidence rates, and all the numbers were based on the same, real data. When only 5-year survival rates were presented, 66% of the physicians recommended screening; 78% were convinced of the screening's efficacy and showed the highest overestimation of the screening's benefit. In contrast, when confronted with mortality rates, only 8% gave a recommendation and only 5% considered the screening to be efficient. The two combined versions produced results in between these values. This study illustrates the influence of the format used to describe screening effects on measures of risk reduction as well as behavioral measures. Only annual mortality rates convey a transparent and unbiased picture of actual changes in mortality that could be the result from screening. As the 5-year survival rate is a highly specific medical concept, it is not surprising that patients would be unaware of its misleading potential, but that even doctors are not aware of these problems is worrying.

Lessons learned: Five-year survival rates do not allow us to adequately evaluate health interventions, particularly screenings. In contrast, annual mortality should be used to illustrate effects.

## Graphical Representations

A promising alternative way to present numerical estimates is with graphs. As the saying goes, a picture is worth a thousand words. In the eighteenth century, William Playfair was the first to use bar charts and pie charts to illustrate economic and political data. Later, at the beginning of the twentieth century, the philosopher and economist Otto Neurath proposed symbols to display statistical information. Since then, different formats have been identified and used to communicate risks. The advantages of graphs are manifold (see Lipkus and Hollands 1999; Ancker et al. 2006; Lipkus 2007; Zikmund-Fisher et al. 2008a, b):

1. *Graphs often attract and maintain attention:* Attention and the expectation of successfully handling quantitative information are important prerequisites for a motivated patient to get involved in personal health decisions.
2. *Graphs foster automatic and intuitive processes:* A transparent and well-designed graph reduces the cognitive effort needed to extract and understand the information. For instance, certain formats do some of the mathematical operations for the observer and consequently facilitate understanding (e.g., part-to-whole relationships; ratio concepts).
3. *Data patterns can be detected:* Some graphs help display data over a longer period of time and show patterns and trends in the long run (e.g., lifetime incidence).
4. *Graphs can help communicate uncertainty:* A main challenge in risk communication is the inclusion of uncertainty parameters. Graphs can facilitate displaying uncertainty transparently.
5. *Graphs can overcome low numeracy:* Statistical illiteracy and innumeracy are widespread phenomena. Especially those low in numeracy tend to misunderstand risk information. Graphs offer an alternative format to help people with low numeracy understand risks and make informed decisions.

Although graphs facilitate the communication of statistical information, a variety of visualization formats can be used to provide the same information – and not all are equally effective. Again it is a question of ecologically structured information. Just like numbers, some graphs have the potential to display information in a biased way and consequently mislead the public. Additionally, understanding graphs requires basic skills to extract the relevant information and read beyond the data given – an ability described as graph literacy.

## The Example of Icon Arrays

One format to visually display risks is icon arrays (pictographs). Icon arrays are graphical visualizations consisting of icons (faces, circles, figures) that represent individuals belonging to a certain group or a group as a whole (see Fig. 24.3).

Fagerlin et al. (2005) tested the impact of pictographs on a hypothetical treatment choice in an angina scenario. Subjects had to choose between two treatments and received anecdotal evidence about success and failure rates of the respective treatments. Anecdotal evidence was either representative of the success rates or not. The treatment with the higher success rate was chosen more often when anecdotal evidence was representative in comparison to when it was not. However, when pictographs were included, the representativeness effect of the anecdotes diminished. By including icon arrays, the influence of narrative evidence can be reduced.

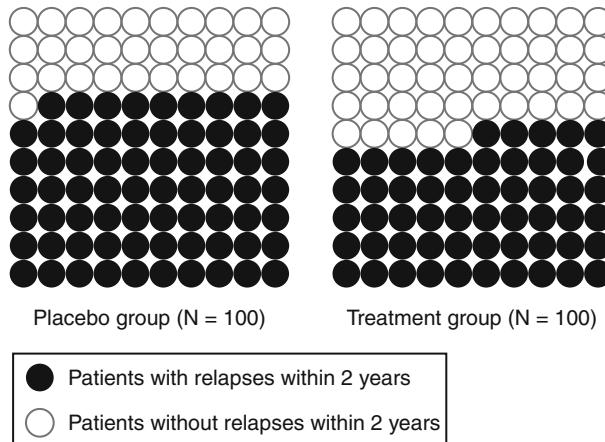


Fig. 24.3

**Icon array: Benefit of interferon therapy for multiple sclerosis patients (modified from Heesen et al. 2008, p. 40).** The treatment aims to reduce multiple sclerosis relapses. Displayed are two groups – a placebo group (*left*) and a treatment group (*right*) – with 100 patients in each. In the placebo group, out of 100 patients 31 patients had no multiple sclerosis relapse within 2 years. In the treatment (interferon therapy) group, 45 had no multiple sclerosis relapse within 2 years. In other words, 14 out of 100 patients benefit from interferon therapy

Icon arrays also help reduce base rate neglect (Garcia-Retamero and Galesic 2009; Garcia-Retamero et al. 2010). Subjects had to compare a treatment group with a control group. When the two groups have equal sample sizes, subjects showed no difficulties in estimating treatment effects. However, the proportion of correct estimates decreased when the groups differed in sample size – the denominator information was often neglected. Icon arrays call attention to the different sample size in the treatment and control group and thereby help to reduce base rate neglect.

It could be argued whether icon arrays facilitate understanding of risk information in addition to transparent numerical formats. But in fact, graphs have an incremental effect on risk perception. In one study, subjects received treatment effects framed as absolute or relative risk reduction with or without icon arrays (Galesic et al. 2009). Icon arrays improved performance for both framing formats. Especially those low in numeracy benefited from the additional visualization.

When Stone et al. (1997) compared a numerical risk reduction format (30–15) with stick figures to understand people's willingness to pay for improved toothpaste versus standard toothpaste, they observed a higher willingness to pay in the graphical condition. They interpreted the result as meaning that graphs led to higher risk avoidance in comparison to numbers. However, while the base rate was presented in the numerical condition, the stick figures did not represent the part–whole relationship. In later experiments (Stone et al. 2003), the stick figures were replaced by charts and stacked bar charts. The inclusion of the part–whole relationship in the graphical formats reversed the original effect: Willingness to pay was lower in the graph conditions than in the numerical conditions. Similarly, Garcia-Retamero and

Galesic (2010) found that bar charts and icon arrays primarily improve risk understanding when the population at risk is included. Pie charts and horizontal bar charts helped people with low numeracy to overcome framing effects.

## Graph Literacy

Despite graphs being more intuitive and facilitating the interpretation of statistical information, their understanding requires basic skills. Galesic and Garcia-Retamero (2011) developed a 13-item scale to measure graph literacy in the medical domain. The scale is meant to differentiate people's ability to read graphical information (e.g., find the correct data), to read between the data (find relationships, e.g., to identify which proportion in a pie chart is larger), and also beyond the data (e.g., inferences and predictions that can be derived from the data). The scale covers different graph formats such as bar graphs, pie charts, line graphs, and icon arrays. The authors validated the scale on probabilistic samples in Germany and the United States and reported correlations between graph literacy and education in Germany (0.29) and the United States (0.54) and between graph literacy and numeracy in Germany (0.47) and the United States (0.55). The majority of people who are low in numeracy are also low in graph literacy and those high in numeracy are usually higher in graph literacy. Primarily those low in numeracy but high in graph literacy benefited from the inclusion of visualizations; those high in numeracy but low in graph literacy had no incremental benefits.

Notwithstanding the advantages of graphs to communicate risks, shortcomings of visual displays cannot be overlooked. People who lack basic graph literacy skills have difficulties extracting and interpreting the relevant information. Up to now, little has been known about how graphical, text, and numerical information interplay. People might shift all their attention to graphs, thereby ignoring any other information relevant to a decision. Additionally, more research is needed to know under what condition a particular visualization format works best. Some graphs are better than others for communicating lifetime prevalence or contrasting different treatment effects. Last but not least, many graphs in newspapers and the scientific literature look fancy but are in fact uninformative or even misleading. Wainer (1984) summarized the 12 most powerful techniques for displaying data badly. In sum, as Lipkus (2007) pointed out, "knowledge of how graphical displays affect risk perceptions is still in its infancy and remains, with few exceptions, a largely atheoretical research area" (p. 702).

Lessons learned: Graphs are a promising tool for improving people's interpretation of risks. However, research about graphs is still in its infancy.

## Uncertainty Communication

The French philosopher Voltaire once said "Doubt is not a pleasant condition, but certainty is absurd." Our world is fundamentally uncertain and yet many people cling to the illusion of certainty. Knight (1921) previously distinguished between risk, which can be computed numerically, and uncertainty, which is immeasurable. A less strict definition of uncertainty allows estimating parameters of uncertainty probabilities, such as standard deviations, confidence intervals, or experts' confidence ratings. A further distinction is made between

uncertainty and variability (Thompson 2002). Variability refers to the fact that different individuals or groups in a population have different risks. This could be due to differences in age, gender, region, or exposure to risk factors. This is different from uncertainty, which refers to imperfect knowledge. Although often ignored, uncertainties play a major role in medical decision making (see, e.g., Politi et al. 2007). First, scientific evidence is limited. Even randomized controlled trials – often regarded as the gold standard in medical research – have limitations due to design principles, sample size, and lack of validity and reliability of measures. Second, risk estimates are based on population data and therefore cannot be applied one-to-one to individuals (see example of individual vs. public health perspective on pap smear screening above). Third, risk estimates are based on past events. Their application to the present and future rests on the assumption that the environment and underlying forces do not change.

### Fear of Disclosing Uncertainty

Uncertainty communication is more the exception than the rule. When in 2009 the World Health Organization proclaimed the H1N1 pandemic, it was assumed that worldwide about two billion people could become infected and between 2 and 7.4 million people could die. This forecast influenced many policy decisions, such as the implementation of H1N1 vaccination programs (Feufel et al. 2010). However, at that point of time, little was known about the actual severity and spread of the virus. In hindsight, this prediction appears exaggerated. Why do experts shrink from disclosing and communicating uncertainty?

First, many experts believe that the public is incapable of understanding uncertainty (Frewer et al. 2003). Communicating uncertainty confuses people who are mainly ambiguity averse and uncertainty intolerant (Epstein 1999). Second, experts might fear losing trust and perceived competence if they reveal that some aspects of an issue are unknown. However, as the H1N1 example illustrates, the opposite can happen: Maintaining an illusion of certainty that in hindsight turns out to be false decreases trust in the expert (Holmes et al. 2009; Feufel et al. 2010). Narrative evidence underlines this conclusion. When the Bank of England started to publish the protocols of their board meetings and transparently reveal related uncertainty in their prognoses about economic growth, the British public rated it as the most trusted institution (Gigerenzer 2007, p. 215).

### How to Communicate Uncertainty

The core question is still how to transparently communicate uncertainty. Ibrekk and Morgan (1987) studied laypeople's understanding of nine different graphical uncertainty representations. Subjects received graphs of weather forecasts about the probability of snow (without any explanations) and the prediction of water depth in a flood (with explanations). As a dependent measure, subjects had to assess the most likely estimate as well as the range. Findings showed that a point estimate including a 95% confidence interval and a Tukey box were easiest to understand. The presence of an explanation led to a slight improvement. Subjects were most sure about their estimates if the point estimate plus confidence interval or a histogram was provided. Cumulative density functions and pie charts seemed improper for communicating risks. This was among the first studies investigating tools to communicate uncertainty to laypeople.

Johnson and Slovic (1995) conducted four studies to understand the influence of uncertainty information. They concluded that people are unfamiliar with uncertainty information and that the recognition of uncertainty caused trouble. However, graphical tools can help communicate uncertainty; the disclosure of uncertainty also signaled honesty for some people but was a sign of incompetence for others. To better understand laypeople's perception of uncertainty, Schapira et al. (2001) formed focus groups and asked women whether an uncertainty statement (in this case a confidence interval) should be included in risk communication. More highly educated women appreciated the inclusion of confidence intervals and interpreted the information as more complete, whereas less educated women reacted with a decrease in trust and the dilution of the actual treatment benefit. Another study investigated effects of doctor's uncertainty disclosure on breast cancer patients (Politi et al. 2010). Although women's breast cancer treatment choice and consistency with expert's opinion was independent of the doctor's disclosure of uncertainty, uncertainty communication reduced decision satisfaction.

With respect to risk perception, it is assumed that increased uncertainty leads to an increase in perceived risks. Put differently, the more uncertain a hazard, the greater the worry that is associated with it, which in turn shifts people's focus to bad outcomes (Einhorn and Hogarth 1985; Viscusi et al. 1991). Yet Kuhn (2000) found that the communication of uncertainty interacts with prior attitude. Subjects were split into two groups defined by high or low environmental concern and received five different scenarios describing environmental hazards. When the risk information was presented as a point estimate, environmental attitudes predicted environmental risk perception. However, the differences between the two groups reduced when an uncertainty statement was included, primarily because those with high environmental concern showed lower perceived risks. A potential explanation is that people with high environmental concern appreciated the uncertainty information (either as a verbal or numerical statement) and perceived the communicator as more honest – similar to the conclusion drawn by Johnson and Slovic (1995). These results are also in line with other proposals, namely, that one way to increase credibility and trust is to present uncertainty instead of maintaining an illusion of certainty (Frewer 1999; van Dijk et al. 2008).

Interestingly, uncertainty communication can improve decisions with respect to a normative criterion (Nadav-Greenberg and Joslyn 2009). Subjects played a road treatment task and had to decide whether to salt roads based on either a point estimate ("It will be 1.7°C tomorrow") or an uncertainty forecast ("It will be 1.7°C tomorrow, but there is an 18% chance that it will freeze"). For every treatment of the road, an amount of \$1,000 had to be paid. However, if the decision maker did not salt the roads and the temperature dropped to freezing, a penalty of \$6,000 had to be paid. According to the expected value, the roads should be treated when the probability of freezing is above 16.7%. Subjects with a deterministic forecast showed a larger deviation from expected value compared to subjects with probabilistic forecasts, resulting in a higher end budget for those who had access to the deterministic forecast *and* probability of a freeze.

In sum, evidence suggests that uncertainty communication is not as prejudicial as often believed. Laypeople understand and even appreciate uncertainty information. There is no evidence that confirms experts' fear of losing trust and perceived credibility when disclosing uncertainty – the opposite effect might be true. However, this is only a tentative conclusion; more research is needed to better understand how uncertainty communication affects and interacts with trust, credibility, and decision-making processes. People have different expectations in different domains that affect their willingness to accept uncertainty – these inter- and

intraindividual differences should be examined to improve risk communication and make uncertainties more salient.

Lessons learned: Uncertainty communication is not bad per se. How it finally affects decision quality, trust, credibility, and satisfaction still requires further examination.

## Further Research

---

Throughout this chapter, we have presented and discussed the current state of research in health risk communication. The research presented here so far has sought to support the idea that transparent risk communication can inform the public and become the basis for shared decision making. In the following, we will discuss what boundaries exist that undermine transparent risk communication and how to overcome them. The following three aspects are at the center of our discussion:

1. Research gaps: Where is further research needed?
2. Political and societal obstacles: Why is risk communication rarely put into practice?
3. Teaching statistical literacy: How can society become statistically literate?

Whereas the first point primarily addresses the research community, the latter two require transferring research into practice and demand the interaction between researchers and policy makers.

### Research Gaps

---

While research in risk communication has already identified multiple formats and alternatives for representing statistical information, there are still many open research questions to be addressed. Here are three potential research fields.

### Individual Differences

It would be interesting to look at the role of interindividual differences in understanding risk communication. People do not only differ with respect to numeracy and graph literacy skills, which have important implications for their understanding of risks. Other frequently discussed factors are age, education, socioeconomic status, intelligence, need for cognition, prior experience, and media competency. Especially the development of patient support technologies, which aim at providing tailored information to individuals, requires knowledge about how individual differences affect information search and decision making, and how to discover these differences. With the help of tailored information, patients can evaluate and select those treatments that fit their personal preferences and needs.

Let us illustrate this with a fictitious example: Imagine three patients who are thinking about participating in colon cancer screening. They have four alternatives: fecal-occult blood test, DNA test, colonoscopy, and sigmoidoscopy. The screening programs differ on various dimensions like how well the test detects early stages of cancer, how often the test errs, how invasive the treatment is, what the side effects are, and how much the test costs. One patient searches for information about all four alternatives on all dimensions. He weights each

question according to his preferences and adds up his evaluations of all treatments. Finally, he selects the one with the highest score. This is called a weighting and adding strategy. However, the second patient follows a different strategy: She thinks that the test should be as good as possible in detecting early stages of cancer. Therefore, she ignores the other questions. If two tests are equally effective, she compares those alternatives on the level of side effects and chooses the one with less severe side effects. Using this heuristic, she might select a different treatment from that chosen by the first patient. A third patient does not search for any information about the treatment and simply asks his doctor for a recommendation. Tailored information needs to be designed to satisfy the consumer's information search and decision strategy. Research about individual differences is still rare but is crucial to understanding the interplay between risk communication and cognitive strategies.

## **Integrating Information Sources**

Another research branch addresses the interaction of different information sources and its effect on health decisions. To learn about health treatments, a patient can consult many different sources: health professionals, friends, patient associations, newspaper reports, and the World Wide Web. Especially the latter provides a new and prominent platform to learn more about health treatments. The Web has both advantages and disadvantages. There are almost no limits on what information patients can search for, and information is often up to date. However, at the same time, patients have to evaluate this information and judge the credibility and trustworthiness of various sources, which is particularly important as even short exposure to misinforming Web sites can have a lasting influence on people's risk perception (Betsch et al. 2010). Patients must decide to what extent they can, for instance, trust information from the pharmaceutical industry or how objective the information is on vaccine-critical Web sites. Evidence and opinions on the Web are often conflicting and might confuse rather than educate the consumer. Again, patients run the risk of being misinformed and misled if they rely on doubtful evidence. Research can help us understand people's Web information search behavior (e.g., Hargittai et al. 2010) and how people identify reliable sources. Also, research has started to identify the relevant skills required to use the Internet successfully (Hargittai 2005, 2009), which could lead to the development of interventions to improve those skills in the future.

In addition, the Web has had a direct impact on the doctor–patient relationship (e.g., Diaz et al. 2002; McMullan 2006). Patients are not “naïve” and uninformed any more but have prior attitudes and expectations when meeting their doctors. On the one hand, the doctor–patient interaction benefits from informed patients. Patients and doctors have a more balanced relationship and need less time since the patient is already informed. On the other hand, doctors might have difficulties disabusing patients of potentially false beliefs and expectations acquired through the Internet. The pros and cons of using the Internet to inform patients and increase expertise in laypeople still needs further research.

## **Implementing Theories of Risk Communication**

Another issue that researchers have to address that relates to the intersection between research and practice is the lack of theoretical frameworks in the field of risk communication, and in

decision aids in particular. Durand et al. (2008) evaluated 50 decision support technologies with respect to their theoretical frameworks. Only 17 were explicitly based on a theoretical framework, mainly expected-utility theory. The lack of theories and their application in the design of decision aids, as well as in the evaluation process, leads to insufficient and ad hoc constructed decision aids. Therefore, researchers need (1) to formulate and test theories on which decision support technologies can be based, (2) to make these theories available to designers of decision support technologies, and (3) to evaluate the implementation of the theories in the decision support technologies.

## Obstacles to Implementing Risk Communication

---

Investigating and identifying transparent risk communication formats is one requirement to improve risk communication in society. However, the second step is to transfer research into practice. As the lack of theories in designing decision support technologies shows, there is a long way to go from theory to practice.

Policy makers still communicate relative instead of absolute risk reductions, pharmaceutical industries promote their interests with misleading statistics, and health professionals themselves have difficulties with numbers. Why is this still the case? An answer is found in the seven sins identified by Gigerenzer and Gray (2011), which is already mentioned in the ➤ [Introduction](#). Three of these sins directly address the issue of transparent risk communication, and were part of this chapter: biased reporting in medical journals, pamphlets, and the media. We also alluded to the lack of statistical literacy in health professionals and will discuss consequences and challenges in the next section. We have also already described another sin: defensive decision making. Doctors often do not prescribe those treatments that seem best but are guided by the desire to minimize potential legal consequences. Finally, biased funding refers to the pharmaceutical industry often sponsoring research trials. Consequently, researchers are not free in the research topics they select, the study design, data analysis, and data interpretation since industrial interests need to be considered. One might argue that researchers have to disclose conflicts of interests. However, this is not always the case. Weinfurt et al. (2008) found that consistent disclosure of financial interests was the exception in the biomedical literature and asked editors and authors to take responsibility.

One movement that has tried to eliminate industrial interests in research trials was launched by *JAMA*, the *Journal of the American Medical Association*. The editors launched a requirement for independent statistical analysis of industry-driven research in 2005. In contrast to other medical journals, *JAMA* published fewer RCTs and also fewer industry-driven RCTs (Wager et al. 2010). Further advances have been made by the introduction of standards and guidelines for reporting observational research (i.e., STROBE statement 2007; CONSORT statement 2009). Simple checklists help authors and editors evaluate research reports and assure complete and unbiased reporting. Such guidelines and standards should not be restricted to scientific journals but should also be set for media health coverage and health advertisements.

Media analyses and advertisement content analyses have repeatedly shown that the media rarely communicate numbers, and when they do, they use biased formats. Yet the mass media have the power to shape health decisions (Grilli et al. 2009) and thereby intentionally or

unintentionally misinform and mislead the public. Health promotion campaigns from the pharmaceutical industry primarily follow financial interests and persuade rather than inform. For instance, in the United States in 2008 the pharmaceutical industry ranked second (behind the automotive industry) in dollars spent on advertising (Nielsen 2009).

An alternative approach to advertising is the use of so-called facts boxes (Schwartz et al. 2007). Facts boxes summarize the current state of evidence about drugs or other treatments in a way that laypeople can easily understand. They cover basic information and provide numbers in transparent formats by contrasting treatment and placebo groups and hence serve an educational purpose. ↗ *Figure 24.1* represents a facts box with basic information about mammography screening based on current scientific evidence.

## Teaching Statistical Literacy

---

Many people in our society are statistically illiterate and innumerate. This phenomenon applies not only to laypeople, but also to experts. A way out of this dilemma is to promote education in statistical thinking on at least three levels: Statistics should be taught in schools, and statistics training should be offered to health professionals and (science) journalists.

### Statistical Teaching in Schools

Statistical thinking is hardly taught at schools. Mathematics curricula do not include teaching statistical concepts; instead, the focus is on the mathematics of certainty, such as algebra, geometry, and trigonometry. In contrast to a widespread belief that children cannot deal with statistics, children at the elementary school level are already capable of understanding fundamental concepts of statistical thinking, such as natural frequencies and icon arrays (Zhu and Gigerenzer 2006). Hands-on approaches to problem solving, such as with tinker cubes, lego-like units, allow even first graders to learn about conditional frequencies through play (Kurz-Milcke and Martignon 2007; Kurz-Milcke et al. 2008). Despite attempts to include statistics in school curricula, there are four constraints that undermine successful and sustainable implementation. First, the first contact with statistics occurs too late in schools. Second, many textbooks use confusing representation formats. Third, statistics are often taught in a pallid way by abstract and unrealistic examples. Fourth, teachers themselves are often not as familiar with these concepts as they ought to be. A rethinking in mathematical teaching is pivotal for future statistically literate generations.

### Statistics Training Education for Health Professionals

The second step addresses education of health professionals. Doctors directly interact with patients and therefore require the skills not only to understand statistics, but also to transparently communicate them. As far back as 1937 an editorial in the Lancet called attention to the strong link between medicine and statistics and the lack of fundamental abilities of doctors to deal with statistical information (“Mathematics and Medicine” 1937). It stated that the use

(or abuse) of statistics “tends to induce a strong emotional reaction in non-mathematical minds.” It complained that for “most of us figures impinge on an educational blind spot,” which “is a misfortune, because simple statistical methods concern us far more closely than many of the things that we are forced to learn in the 6 long years of the medical curriculum.” What has changed since then? Doctors still have trouble calculating positive predictive values and are prone to framing effects (e.g., 5-year survival rates vs. mortality rates, or relative vs. absolute risk reduction). Health professionals need to be trained in statistics. This will teach them how to identify biased reporting and how to translate statistical information into transparent formats.

### Statistics Education for (Science) Journalists

The third target population is journalists. As previously mentioned, the mass media play an important role in educating the public. However, journalists might just reproduce biased reporting that has its origin in the medical literature. Therefore, educating scientists and making them aware of these biases will help them see through embellishments and obfuscations to translate risk information into comprehensive formats. They may also put public pressure on those who practice biased reporting.

## Conclusion

---

Risk communication is a requirement for an informed public to be able to adequately deal with risks and uncertainties. On the one hand, experts and laypeople have difficulties in dealing with statistical information. On the other hand, the problem is less in people’s minds and more in a health environment that puts little effort into presenting risks in an unbiased way. Biased reporting encompasses the omission of important information as well as the use of nontransparent communication formats. Informed and shared decision making will remain an illusion unless transparent risk communication formats are consistently applied.

We believe that statistically literate patients improve health decisions on an individual as well as on a public health level. Throughout this chapter, we proposed ways to design risk communication to educate and inform patients, instead of persuading them. These points should be kept in mind (see  [Table 24.3](#)):

- Absolute risk changes are preferred over relative risk changes.
- Natural frequencies facilitate Bayesian reasoning in comparison to conditional probabilities.
- Annual mortality rates are less misleading and less biased than 5-year survival rates.
- Graphs can help overcome innumeracy.
- Disclosing uncertainty can help overcome the illusion of certainty.

People are able to make personal decisions that reflect their preferences and needs when they have sufficient information on which to base their decisions. There are two fundamental “adjustment screws”: the consequent application of transparent communication formats and the implementation of education programs on different societal levels. Last but not least,

**Table 24.3**

**Nontransparent versus transparent communication of risks: Four examples of how risks can be communicated to mislead and misinform the public and their transparent counterparts**

How to communicate risks nontransparently	How to communicate risks transparently
<p><i>Relative risks</i> “The new generation of the contraceptive pill increases the risk of thrombosis by 100%.”</p>	<p><i>Absolute risks</i> “The new generation of the contraceptive pill increases the risk of thrombosis from 1 in 7,000 to 2 in 7,000.”</p>
<p><i>Conditional probabilities</i>            – The probability of breast cancer is 1% for a woman at age 40 who participates in routine screening (this is the prevalence or base rate)            – If a woman has breast cancer, the probability is 90% that she will get a positive mammography (this is the sensitivity or hit rate)            – If a woman does not have breast cancer, the probability is 9% that she will also get a positive mammography (this is the false-positive rate)            What is the probability that a woman at age 40 who had a positive mammogram actually has breast cancer?  <math>P(H D) = \frac{0.9 \cdot 0.01}{0.9 \cdot 0.01 + 0.09 \cdot 0.99} = 0.092</math> </p>	<p><i>Natural frequencies</i>            – Ten out of 1,000 women at age 40 who participate in mammography screening have breast cancer (prevalence or base rate)            – Of these 10 women, 9 have a positive mammogram (sensitivity or hit rate)            – Out of the 990 women who do not have breast cancer, about 89 will have a positive mammogram nonetheless (false-positive rate)            Now imagine a representative sample of 1,000 women age 40 who participate in breast cancer screening. How many of these women with a positive test result actually have breast cancer?  <math>P(H D) = \frac{9}{9+89} = 9.2</math> </p>
<p><i>Five-year survival rate</i> “The 5-year survival rate for people diagnosed with prostate cancer is 98% in the USA vs. 71% in Britain.”</p>	<p><i>Annual mortality rate</i> “There are 26 prostate cancer deaths per 100,000 American men versus 27 per 100,000 men in Britain.”</p>
<p><i>Single-event probability</i> “If you take Prozac, the probability that you will experience sexual problems is 30–50% (or: 30 to 50 chances out of 100).”</p>	<p><i>Frequency statement</i> “Out of every 10 of my patients who take Prozac, 3–5 experience a sexual problem.”</p>

lessons learned in health risk communication can be adapted to other domains as well. Transparency and statistical literacy help people evaluate financial, environmental, and technological risks, and enable society to competently meet future challenges.

## References

- Allen M, Preiss R (1997) Comparing the persuasiveness of narrative and statistical evidence using meta-analysis. *Commun Res Rep* 14:125–131
- Ancker JS, Kaufman D (2007) Rethinking health numeracy: a multidisciplinary literature review. *J Am Med Inform Assoc* 14:713–721
- Ancker JS, Senathirajah Y, Kukafka R, Starren JB (2006) Design features of graphs in health risk communication: a systematic review. *J Am Med Inform Assoc* 3:608–618
- Baesler JE (1997) Persuasive effects of story and statistical evidence. *Argument Advocacy* 33:170–175

- Baesler JE, Burgoon JK (1994) The temporal effects of story and statistical evidence on belief change. *Commun Res* 21:582–602
- Barbey AK, Sloman SA (2007) Base-rate respect: from ecological rationality to dual processes. *Behav Brain Sci* 30:241–297
- Barton A, Mousavi S, Stevens JR (2007) A statistical taxonomy and another “chance” for natural frequencies. *Behav Brain Sci* 30:255–256
- Berry D, Raynor T, Knapp P, Bersellini E (2004) Over the counter medicines and the need for immediate action: a further evaluation of European commission recommended wordings for communicating risk. *Patient Educ Couns* 53:129–134
- Betsch C, Renkewitz F, Betsch T, Ulshöfer C (2010) The influence of vaccine-critical Internet pages on perception of vaccination risks. *J Health Psychol* 15:446–455
- Bodemer N, Müller SM, Okan Y, Garcia-Retamero R, Neumeyer-Gromen A (2011) Do the media provide transparent health information? A cross-cultural comparison of public information about the HPV vaccine (Submitted)
- Brase GL (2009) Pictorial representations in statistical reasoning. *Appl Cogn Psychol* 23:369–381
- Brun W, Teigen KH (1988) Verbal probabilities: ambiguous, context-dependent, or both? *Organ Behav Hum Decis Process* 41:390–404
- Bucher HC, Weinbacher M, Gyr K (1994) Influence of method of reporting study results on decision of physicians to prescribe drugs to lower cholesterol concentration. *Br Med J* 309:761–764
- Budescu DV, Wallsten TS (1985) Consistency in interpretation of probabilistic phrases. *Organ Behav Hum Decis Process* 36:391–405
- Casscells W, Schoenberger A, Grayboys T (1978) Interpretation by physicians of clinical laboratory results. *N Engl J Med* 299:999–1000
- Consort (2009) Consolidated standards of reporting trials. [www.consort-statement.org](http://www.consort-statement.org). Accessed April 2011
- Cosmides L, Tooby J (1996) Are humans good intuitive statisticians after all? Rethinking some conclusions of the literature on judgment under uncertainty. *Cognition* 58:1–73
- Covey J (2007) A meta-analysis of the effects of presenting treatment benefits in different formats. *Med Decis Making* 27:638–654
- Davids SL, Schapira MM, McAuliffe TL, Nattinger AB (2004) Predictors of pessimistic breast cancer risk perception in a primary care population. *J Gen Intern Med* 19:310–315
- deWit JBF, Das E, Vet R (2008) What works best: objective statistics or a personal testimonial? An assessment of the persuasive effects of different types of message evidence on risk perception. *Health Psychol* 27:110–115
- Diaz JA, Griffith RA, Ng JJ, Reinert SE, Friedmann PD, Moulton AW (2002) Patients’ use of the Internet for medical information. *J Gen Intern Med* 17:180–185
- Dieckmann NF, Slovic P, Peters E (2009) The use of narrative evidence and explicit likelihood by decisionmakers varying in numeracy. *Risk Anal* 29:1473–1488
- Djulbegovic M, Beyth RJ, Neuberger MM, Stoffs TL, Vieweg J, Djulbegovic B et al (2010) Screening for prostate cancer: systematic review and meta-analysis of randomised controlled trials. *Br Med J* 341:c4543
- Dören M, Gerhardus A, Gerlach FM, Hornberg C, Kochen MM, Kolip P et al. (2008) Wissenschaftler/innen fordern Neubewertung der HPV-Impfung und ein Ende der irreführenden Informationen. [http://www.uni-bielefeld.de/gesundhw/ag3/downloads/Stellungnahme\\_Wirksamkeit HPV-Impfung.pdf](http://www.uni-bielefeld.de/gesundhw/ag3/downloads/Stellungnahme_Wirksamkeit HPV-Impfung.pdf). Accessed April 2011
- Durand M-A, Stiel M, Boivin J, Elwyn G (2008) Where is the theory? Evaluating the theoretical frameworks described in decision support technologies. *Patient Educ Couns* 71:125–135
- Eddy DM (1982) Probabilistic reasoning in clinical medicine: problems and opportunities. In: Kahneman D, Slovic P, Tversky A (eds) *Judgment under uncertainty: heuristics and biases*. Cambridge University Press, Cambridge, pp 246–267
- Edwards A, Elwyn G (2009) Shared decision making in health care: achieving evidence-based patient choice. Oxford University Press, Oxford
- Edwards A, Elwyn G, Covey J, Matthews E, Pill R (2001) Presenting risk information – A review of the effects of “framing” and other manipulations on patient outcomes. *J Health Commun* 6:61–82
- Einhorn HJ, Hogarth RM (1985) Ambiguity and uncertainty in probabilistic inference. *Psychol Rev* 92:433–461
- Epstein LG (1999) A definition of uncertainty aversion. *Rev Econ Stud* 66:579–608
- Erev I, Cohen BL (1990) Verbal versus numerical probabilities: efficiency, biases, and the preference paradox. *Organ Behav Hum Decis Process* 45:1–18
- Estrada CA, Martin-Hryniwicz M, Peek BT, Collins C, Byrd JC (2004) Literacy and numeracy skills and anticoagulation control. *Am J Med Sci* 328:88–93
- European Commission (1998) A guideline on the readability of the label and package leaflet of medicinal products for human use. EC Pharmaceuticals Committee, Brussels
- Fagerlin A, Wang C, Ubel PA (2005) Reducing the influence of anecdotal reasoning on people’s health care decisions: is a picture worth a thousand statistics? *Med Decis Making* 25:398–405
- Fagerlin A, Zikmund-Fisher BJ, Ubel PA, Jankovic A, Derry HA, Smith DM (2007) Measuring numeracy

- without a math test: development of the subjective numeracy scale. *Med Decis Making* 27:672–680
- Feufel M, Antes G, Gigerenzer G (2010) Competence in dealing with uncertainty lessons to learn from the influenza pandemic (H1N1) 2009. *Bundesgesundheitsblatt* 53:1283–1289
- Frewer LJ (1999) Risk perception, social trust, and public participation in strategic decision making: implications for emerging technologies. *Ambio* 28:569–574
- Frewer LJ, Hunt S, Brennan M, Kuznesof S, Ness M, Ritson C (2003) The views of scientific experts on how the public conceptualize uncertainty. *J Risk Res* 6:75–85
- Frosch DL, Krueger PM, Hornik RC, Cronholm PE, Barg FK (2007) Creating demand for prescription drugs: a content analysis of television direct-to-consumer advertising. *Ann Fam Med* 5:6–13
- Gaissmaier W, Straubinger N, Funder DC (2007) Ecologically structured information: the power of pictures and other effective data presentations. *Behav Brain Sci* 30:263–264
- Gal I (1995) Big picture: what does “numeracy” mean? <http://mathforum.org/teachers/adult.ed/articles/gal.html>. Accessed April 2011
- Galesic M, Garcia-Retamero R (2010) Statistical numeracy for health: a cross-cultural comparison with probabilistic national samples. *Arch Intern Med* 170:462–468
- Galesic M, Garcia-Retamero R (2011) Graph literacy: a cross-cultural comparison. *Med Decis Making* 31:444–457
- Galesic M, Garcia-Retamero R, Gigerenzer G (2009) Using icon arrays to communicate medical risks: overcoming low numeracy. *Health Psychol* 28:210–216
- Garcia-Retamero R, Galesic M (2009) Communicating treatment risk reduction to people with low numeracy skills: a cross-cultural comparison. *Am J Public Health* 99:2196–2202
- Garcia-Retamero R, Galesic M (2010) Who profits from visual aids: overcoming challenges in people’s understanding of risks. *Soc Sci Med* 70:1019–1025
- Garcia-Retamero R, Galesic M, Gigerenzer G (2010) Do icon arrays help reduce denominator neglect? *Med Decis Making* 30:672–684
- Gigerenzer G (2002) Calculated risks: how to know when numbers deceive you. Simon & Schuster, New York
- Gigerenzer G (2007) Gut feeling: the intelligence of the unconscious. Viking, New York
- Gigerenzer G, Gray M (2011) Launching the century of the patient. In: Gigerenzer G, Gray M (eds) Better doctors, better patients, better decisions: envisioning health care 2020. MIT Press, Cambridge, pp 3–28
- Gigerenzer G, Hoffrage U (1995) How to improve Bayesian reasoning without instruction: frequency formats. *Psychol Rev* 102:684–704
- Gigerenzer G, Hoffrage U, Ebert A (1998) AIDS counseling for low-risk clients. *AIDS Care* 10:197–211
- Gigerenzer G, Hertwig R, van den Broek E, Fasolo B, Katsikopoulos KV (2005) “A 30% chance of rain tomorrow”: how does the public understand probabilistic weather forecasts? *Risk Anal* 25:623–629
- Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S (2007) Helping doctors and patients to make sense of health statistics. *Psychol Sci Public Interest* 8:53–96
- Gigerenzer G, Mata J, Frank R (2009) Public knowledge of benefits of breast and prostate cancer screening in Europe. *J Natl Cancer Inst* 101:1216–1220
- Gigerenzer G, Wegwarth O, Feufel M (2010) Misleading communication of risk: editors should enforce transparent reporting in abstracts. *Br Med J* 341:c4830
- Golbeck AL, Ahlers-Schmidt CR, Paschal AM, Dismuke S (2005) A definition and operational framework for health numeracy. *Am J Prev Med* 29:375–376
- Grigg W, Donahue P, Dion G (2007) The nation’s report card: 12th-grade reading and mathematics 2005 (NCES Report No. 2007-468). U.S. Department of Education, National Center for Education Statistics, Washington, DC
- Grilli R, Ramsey C, Minozzi S (2009) Mass media interventions: effects on health care utilization. *Cochrane Database Syst Rev* 1:CD000389
- Gurman AD, Baron J, Armstrong K (2004) The effect of numerical statements of risk on trust and comfort with hypothetical physician risk communication. *Med Decis Making* 24:265–271
- Hacking I (1975) The emergence of probability. Cambridge University Press, Cambridge
- Hargittai E (2005) Survey measures of web-oriented digital literacy. *Soc Sci Comput Rev* 23:371–379
- Hargittai E (2009) An update on survey measures of web-oriented digital literacy. *Soc Sci Comput Rev* 27:130–137
- Hargittai E, Fullerton F, Menchen-Trevino E, Thomas K (2010) Trust online: young adults’ evaluation of web content. *Int J Commun* 4:468–494
- Heesen C, Köpke S, Kasper J, Richter T, Beier M, Mühlhauser I (2008) Immuntherapien der Multiplen Sklerose. [http://www.zmnh.uni-hamburg.de/martin/dl/pinfo/immuntherapien\\_ms\\_inims\\_hamburg.pdf](http://www.zmnh.uni-hamburg.de/martin/dl/pinfo/immuntherapien_ms_inims_hamburg.pdf). Accessed April 2011
- Heesen C, Kleiter I, Nguyen F, Schäffler N, Kasper J, Köpke S et al (2010) Risk perceptions in natalizumab-treated multiple sclerosis patients and their neurologists. *Mult Scler* 16:1507–1512
- Hembroff LA, Holmes-Rovner M, Wills CE (2004) Treatment decision-making and the form of risk

- communication: results of a factorial survey. *BMC Med Inform Decis Mak* 4:1184–1120
- Hoffrage U, Gigerenzer G, Krauss S, Martignon L (2002) Representation facilitates reasoning: what natural frequencies are and what they are not. *Cognition* 84:343–352
- Holmes BJ, Henrich N, Hancock S, Lestou V (2009) Communicating with the public during health crises: experts' experiences and opinions. *J Risk Res* 12:793–807
- Ibrekk H, Morgan MG (1987) Graphical communication of uncertain quantities to nontechnical people. *Risk Anal* 7:519–529
- Johnson BB, Slovic P (1995) Presenting uncertainty in health risk assessment: initial studies of its effects on risk perception and trust. *Risk Anal* 15: 485–494
- Kahneman D, Tversky A (1972) Subjective probability: a judgment of representativeness. *Cognit Psychol* 3:430–454
- Kirsch IS, Jungeblut A, Jenkins L, Kolstad A (2007) Adult literacy in America: a first look at the findings of the National Adult Literacy Survey (NCES Report No. 1993–275; 3rd edn, U.S. Department of Education, National Center for Education Statistics. Washington, DC
- Knapp P, Raynor DK, Berry DC (2004) Comparison of two methods of presenting risk information to patients about the side effects of medicines. *Qual Saf Health Care* 13:176–180
- Knight FH (1921) Risk, uncertainty and profit. Harper, New York
- Kotler P, Lee NR (2007) Social marketing: influencing behaviors for good. Sage, London
- Kuhn KM (2000) Message format and audience values: interactive effects of uncertainty information and environmental attitudes on perceived risk. *J Environ Psychol* 20:41–51
- Kurzenhäuser S (2003) Welche Informationen vermitteln deutsche Gesundheitsbroschüren über die Screening-Mammographie? *Z Arztl Fortbild Qualitatssich* 97:53–57
- Kurz-Milcke E, Martignon L (2007) Stochastische Urnen und Modelle in der Grundschule (Stochastic urns and models in elementary school). In: Kaiser G (ed) Tagungsband der Jahrestagung der Gesellschaft für Didaktik der Mathematik. Verlag Franzbecker, Berlin
- Kurz-Milcke E, Gigerenzer G, Martignon L (2008) Transparency in risk communication: graphical and analog tools. In: Tucker WT, Ferson S, Finkel A, Long TF, Slavin D, Wright P (eds) Strategies for risk communication: evolution, evidence, experience, vol 1128, Annals of the New York Academy of Sciences. Blackwell, New York, pp 18–28
- Kutner M, Greenberg E, Jin Y, Paulsen C (2006) The health literacy of America's adults: results from the 2003 national assessment of adult literacy (NCES Report No. 2006-483). Government Printing Office, Washington, DC
- Lafata JE, Simpkins J, Lamerato L, Poisson L, Divine G, Johnson CC (2004) The economic impact of false-positive cancer screens. *Cancer Epidemiol Biomarkers Prev* 13:2126–2132
- Lewis C, Keren G (1999) On the difficulties underlying Bayesian reasoning: comment on Gigerenzer and Hoffrage. *Psychol Rev* 106:411–416
- Lipkus IM (2007) Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations. *Med Decis Making* 27:696–713
- Lipkus IM, Hollands JG (1999) The visual communication of risks. *J Natl Cancer Inst Monogr* 25:149–162
- Lipkus IM, Peters E (2009) Understanding the role of numeracy in health: proposed theoretical framework and practical insight. *Health Educ Behav* 36:1065–1081
- Lipkus IM, Samsa G, Rimer BK (2001) General performance on a numeracy scale among highly educated samples. *Med Decis Making* 21:37–44
- Macchi L, Mosconi G (1998) Computational features vs frequentist phrasing in the base-rate fallacy. *Swiss J Psychol* 57:79–85
- Malenka DJ, Baron JA, Johansen S, Wahrenberger JW, Ross JM (1993) The framing effect of relative versus absolute risk. *J Gen Intern Med* 8:543–548
- Marcus PM, Bergstrahl EJ, Zweig MH, Harris A, Offord KP, Fontana RS (2006) Extended lung cancer incidence follow-up in the Mayo lung project and over-diagnosis. *J Natl Cancer Inst* 98:748–756
- Marteau TM, Saidi G, Goodburn S, Lawton J, Michie S, Bobrow M (2000) Numbers or words? A randomized controlled trial of presenting screen negative results to pregnant women. *Prenat Diagn* 20:714–718
- Mathematics and medicine (1937, January 2). *Lancet* i:31
- Mazur DJ, Hickham DH, Mazur MD (1999) How patients' preferences for risk communication influence treatment choice in a case of high risk and high therapeutic uncertainty: asymptotic localized prostate cancer. *Med Decis Making* 19: 394–398
- McMullan M (2006) Patients using the Internet to obtain health information: how this affects the patient–health professional relationship. *Patient Educ Couns* 63:24–28
- Moxey A, O'Connell D, McGettigan P, Henry D (2003) Describing treatment effects to patients: how they are expressed makes a difference. *J Gen Intern Med* 18:948–959

- Moynihan R, Bero L, Ross-Degnan D, Henry D, Lee K, Watkins J et al (2000) Coverage by the news media of the benefits and risks of medications. *New Engl J Med* 342:1645–1650
- Mühlhauser I, Kasper J, Meyer G (2006) FEND: understanding of diabetes prevention studies: questionnaire survey of professionals in diabetes care. *Diabetologia* 49:1742–1746
- Nadav-Greenberg L, Joslyn S (2009) Uncertainty forecasts improve decision making among nonexperts. *J Cogn Eng Decis Making* 3:209–227
- Naylor CD, Chen E, Strauss B (1992) Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? *Ann Intern Med* 171:916–921
- Neumeyer-Gromen A, Bodemer N, Müller SM, Gigerenzer G (in press) Ermöglichen Medienberichte und Informationsbroschüren zur Gebärmutterhalskrebsprävention informierte Entscheidungen? Eine Medienanalyse in Deutschland (Submitted)
- Nielsen Company (2009) U.S. ad spending fell 2.6% in 2008, Nielson reports. Nielson Company, New York
- Nisbett RE, Ross LD (1980) Human inference: strategies and shortcomings of social judgment. Prentice-Hall, Englewood Cliffs
- Paulos JA (1988) Innumeracy: mathematical illiteracy and its consequences. Hill & Wang, New York
- Pepper S, Prytulak LS (1974) Sometimes frequently means seldom: context effects in the interpretation of quantitative expressions. *J Res Pers* 8:95–101
- Peters E, Västfjäll D, Slovic P, Mertz CK, Mazzocco K, Dickert S (2006) Numeracy and decision making. *Psychol Sci Public Interest* 17:407–413
- Politi MC, Han PKJ, Col NF (2007) Communicating the uncertainty of harms and benefits of medical interventions. *Med Decis Making* 27:681–695
- Politi MC, Clark MA, Ombao H, Dizon D, Elwyn G (2010) Communicating uncertainty can lead to less decision satisfaction: a necessary cost of involving patients in shared decision making? *Health Expect* 14:1–8
- Reinard JC (1988) The empirical study of the persuasive effects of evidence: the status after fifty years of research. *Hum Commun Res* 15:3–59
- Reyna V, Nelson WL, Han PK, Dieckmann NF (2009) How numeracy influences risk comprehension and medical decision making. *Psychol Bull* 135: 943–973
- Rothman RL, Housam R, Weiss H, Davis D, Gregory R, Gebretsadik T, Shintani A, Elasy TA (2006) Patient understanding of food labels: the role of literacy and numeracy. *Am J Prev Med* 31:391–398
- Sandblom G, Varenhorst E, Rosell J, Löfman O, Carlsson P (2011) Randomised prostate cancer screening trial: 20 year follow-up. *Br Med J* 342:d1539
- Sarfati D, Howden-Chapman P, Woodward A, Salmond C (1998) Does the frame affect the picture? A study into how attitudes to screening for cancer are affected by the way benefits are expressed. *J Med Screen* 5:137–140
- Schapira MM, Nattinger AB, McHorney CA (2001) Frequency or probability? A qualitative study of risk communication formats used in health care. *Med Decis Making* 21:459–467
- Schröder FH, Hugosson J, Roobol MJ et al (2009) Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med* 360: 1320–1328
- Schwartz LM, Woloshin S, Black WC, Welch HG (1997) The role of numeracy in understanding the benefit of screening mammography. *Ann Intern Med* 127:966–972
- Schwartz LM, Woloshin S, Dvorin EL, Welch HG (2006) Ratio measures in leading medical journals: structured review of accessibility of underlying absolute risks. *Br Med J* 333:1248–1252
- Schwartz LM, Woloshin S, Welch HG (2007) Using a drug facts box to communicate drug benefits and harms. *Ann Intern Med* 150:516–527
- Sedrakyan A, Shih C (2007) Improving depiction of benefits and harms: analyses of studies of well-known therapeutics and review of high-impact medical journals. *Med Care* 45:523–528
- Shaw NJ, Dear PRF (1990) How do parents of babies interpret qualitative expressions of probability. *Arch Dis Child* 65:520–523
- Sheridan S, Pignone MP, Lewis CL (2003) A randomized comparison of patients' understanding of number needed to treat and other common risk reduction formats. *J Gen Intern Med* 18:884–892
- Smeeth L, Haines A, Ebrahim S (1999) Numbers needed to treat derived from meta-analyses – sometimes informative, usually misleading. *Br Med J* 318:1548–1551
- Steckelberg A, Balgenorth A, Mühlhauser I (2001) Analyse von deutschsprachigen Verbraucher-Informationsbroschüren zum Screening auf kolorektalem Karzinom. *Z Arztl Fortbild Qualitatsseich* 95:535–538
- Steckelberg A, Berger B, Köpke S, Heesen C, Mühlhauser I (2005) Criteria for evidence-based patient information. *Z Arztl Fortbild Qualitatsseich* 99:343–351
- Steurer J, Held U, Schmidt M, Gigerenzer G, Tag B, Bachmann L (2009) Legal concerns trigger prostate-specific antigen testing. *J Eval Clin Pract* 15:390–392
- Stone ER, Yates JF, Parker AM (1997) Effects of numerical and graphical displays on professed risk-taking behavior. *J Exp Psychol Appl* 3:243–256
- Stone ER, Sieck WR, Bull BE, Yates JF, Parks SC, Rush CJ (2003) Foreground:background salience: explaining

- the effects of graphical displays on risk avoidance. *Organ Behav Hum Decis Process* 90:19–36
- Strobe-Statement (2007) Strengthening the reporting of observational studies in epidemiology. <http://www.strobe-statement.org/>. Accessed April 2011
- Taylor SE, Thompson SC (1982) Stalking the elusive “vividness” effect. *Psychol Rev* 89:155–181
- Thompson KM (2002) Variability and uncertainty meet risk management and risk communication. *Risk Anal* 22:647–654
- Ubel PA, Jepson C, Baron J (2001) The inclusion of patient testimonials in decision aids: effects on treatment choices. *Med Decis Making* 21:60–68
- van Dijk H, Houghton J, van Kleef E, van der Lans I, Rowe G, Frewer LJ (2008) Consumer responses to communication about food risk management. *Appetite* 50:340–352
- Viscusi WK, Magat WA, Huber J (1991) Communication of ambiguous risk information. *Theory Decis* 31:159–173
- Wager E, Mhaskar R, Warburton S, Djulbegovic B (2010) JAMA published fewer industry-funded studies after introducing a requirement for independent statistical analysis. *PLoS One* 5:e13591
- Wainer H (1984) How to display data badly. *Am Stat* 38:137–147
- Wallsten TS, Fillenbaum S, Cox JA (1986) Base rate effects on the interpretation of probability and frequency expressions. *J Mem Lang* 25:571–587
- Wallsten TS, Budescu DV, Zwick R, Kemp SM (1993) Preferences and reasons for communicating probabilistic information in verbal or numerical terms. *Bull Psychon Soc* 31:135–138
- Weber EU, Hilton DJ (1990) Contextual effects in the interpretations of probability words: perceived base rate and severity of events. *J Exp Psychol Hum Percept Perform* 16:781–789
- Wegwarth O, Gaissmaier W, Gigerenzer G (2011) Deceiving and informing: the risky business of risk perception. *Med Decis Making* 31:378–379
- Weinfurt KP, Seils DM, Tzeng JP, Lin L, Schulman KA, Califff RM (2008) Consistency of financial interest disclosures in the biomedical literature: the case of coronary stents. *PLoS One* 3:e2128
- Welch HG, Schwartz LM, Woloshin S (2000) Are increasing 5-year survival rates evidence of success against cancer? *J Am Med Assoc* 283:2975–2978
- Wilde J (2009) PSA screening cuts deaths by 20%, says world's largest prostate cancer study. ERSPC press office, carver wilde communications. <http://www.espc-media.org/release090318.php>. Accessed April 2011
- Witte K, Allen M (2000) A meta-analysis of fear appeals: implications for effective public health campaigns. *Health Educ Behav* 27:591–615
- Woloshin S, Schwartz LM, Black WC, Welch HG (1999) Women's perceptions of breast cancer risk: how you ask matters. *Med Decis Making* 19:221–229
- Zhu L, Gigerenzer G (2006) Children can solve Bayesian problems: the role of representation in mental computation. *Cognition* 98:287–308
- Zikmund-Fisher BJ, Smith DM, Ubel PA, Fagerlin A (2007) Validation of the subjective numeracy scale (SNS): effects of low numeracy on comprehension of risk communications and utility elicitation. *Med Decis Making* 27:663–671
- Zikmund-Fisher BJ, Fagerlin A, Ubel PA (2008a) Improving understanding of adjuvant therapy options by using simpler risk graphics. *Cancer* 113:3382–3390
- Zikmund-Fisher BJ, Ubel PA, Smith DM, Derry HA, McClure JB, Stark AT et al (2008b) Communicating side effect risks in a tamoxifen prophylaxis decision aid: the debiasing influence of pictographs. *Patient Educ Couns* 73:209–214
- Zimmer AC (1983) Verbal vs. numerical processing of subjective probabilities. In: Scholz RW (ed) *Decision making under uncertainty*. Elsevier, Amsterdam, pp 159–182

# **25 Risk Perception and Societal Response**

*Lennart Sjöberg*

Stockholm School of Economics, Stockholm, Sweden

Norwegian University of Science and Technology, Trondheim, Norway

<i>Introduction</i> .....	662
<i>Received Models of Risk Perception</i> .....	662
<i>Beyond the Psychometric Model</i> .....	665
<i>Affect and Emotion</i> .....	666
<i>Trust and Antagonism</i> .....	668
<i>Risk and Policy</i> .....	669
<i>Two Applications</i> .....	671
<i>Further Research</i> .....	672
<i>Conclusions</i> .....	672

**Abstract:** Risk perception is important in policy making. Most research on risk perception has been carried out with nonexperts and members of the public at large, but there are some interesting exceptions, notably the study of experts. Very large differences in risk perception usually appear between experts and nonexperts, but they seem to be partly related to responsibility and social validation and not only to knowledge. Models of risk perception have usually been based on Cultural Theory or the Psychometric Model, but they have had only limited success in accounting for perceived risk. The chapter discusses factors which can improve on the explanatory power of risk perception models, such as Interfering with Nature, Risk Sensitivity, and Risk Target (self or others). Emotions and values have also been investigated. Emotions do play an important role in risk perception, but values have so far not been found to be important. “Affect” is an unclear term since it can refer both to emotions and attitudes. Trust has been another focus of research on risk perception. Trust has almost always been conceived as social trust, i.e., trust in people or organizations. Trust in this sense has a limited influence on risk perception. Epistemic trust, i.e., trust in the science behind risk assessments and risk management, is possibly more important than social trust; at any rate, both types of trust should be considered. Finally, new risks appear all the time, and they require new concepts if we are to understand how people perceive and react to them.

## Introduction

---

The field of policy making with regard to risk issues is difficult to navigate because members of the public often have strong views which differ dramatically from those of experts and administrators, captains of industry, and politicians. Risk issues are often very contentious, and parties tend to have distorted views of each others’ beliefs and values (Sjöberg 1980). In a study of nuclear-waste risk perception (Sjöberg et al. 2000b), it was found that experts and members of the public made systematic errors in estimating each others’ level of perceived risk for a number of risk aspects, see [Figs. 25.1](#) and [25.2](#). Experts made lower risk estimates than the public thought they did. Members of the public made higher risk estimates than the experts thought they did. Hence, the gap between experts and the public was wider than either party believed it to be.

In a democracy, there is a need to understand people’s attitudes and to accommodate policies with those attitudes. Research on risk perception and related attitudes is an important part of this work. Scientific risk estimates are of course important in policy making, but risk perceptions cannot be ignored (Sjöberg 2001b, c). For an excellent review of risk perception research, see Breakwell (2007).

## Received Models of Risk Perception

---

The two best-known models of risk perception are the Psychometric Model (Fischhoff et al. 1978) and Cultural Theory (Douglas and Wildavsky 1982). They have both had a historically important influence on the field, but current work throws doubt on their validity. I summarized my critical analysis of these two models in Sjöberg (2002b), while another paper (Sjöberg 1997) is devoted particularly to Cultural Theory. “Heuristics and biases” (Tversky and Kahneman 1974; Sjöberg 1979) have also played an important role in discussions of risk perception, particularly the availability heuristic (Tversky and Kahneman 1973).

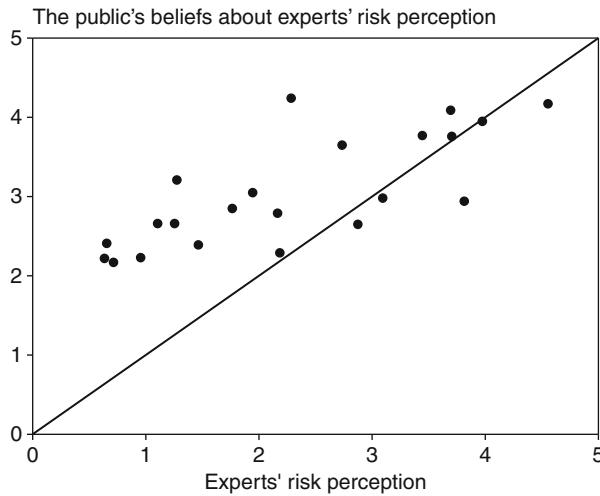


Fig. 25.1

Risk aspects of nuclear waste: experts' risk perceptions and the public's beliefs about the experts' perceptions. Scales refer to size of perceived risk, where 0 = no risk and 5 = a large risk

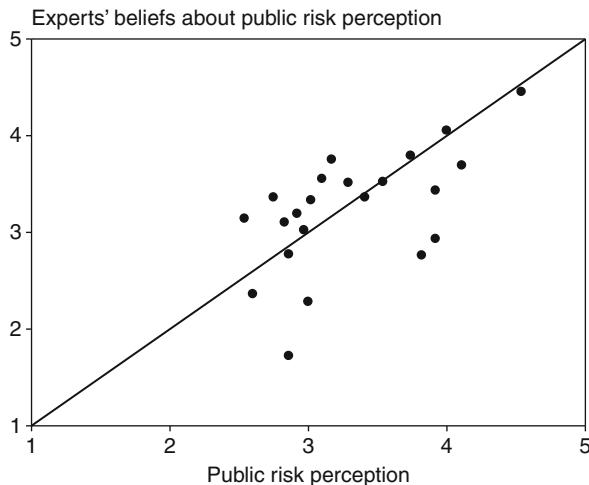


Fig. 25.2

Risk aspects of nuclear waste: the public's risk perceptions and experts' beliefs about the public's perceptions. Scales refer to size of perceived risk, where 0 = no risk and 5 = a large risk

The "heuristics and biases" approach to risk perception stresses availability as a major factor driving perceived risk (Combs and Slovic 1979; McCombs and Gilbert 1986). Information which is more easily brought to mind, more dramatic and striking, etc., is said to be evoke stronger perceptions of risk. Although this principle sounds reasonable, it has not been generally supported by data. In an experimental study, we showed dramatic films about a fire and a nuclear power plant accident and found no tendency toward increased risk perception related to the

contents of the films (Sjöberg and Engelberg 2010). Reasons for perceived risk may be due to the semantics of the concepts instead. For example, “nuclear waste” is a combination of two negatively loaded and threatening concepts, and it is no coincidence that opponents to siting projects make it even more negative by using such terms “atomic garbage.” Industry seems, on the other hand, to prefer the term “spent nuclear fuel.” Brehmer found that people judged road crossings which looked dangerous as risky, even if they had not been exposed to any information about accidents having taken place there (Brehmer 1987).

The basic thrust of the criticism of the Psychometric Model and Cultural Theory is that both models are empirically weak when it comes to accounting for perceived risk of technologies. Scales of Cultural Theory dimensions were suggested first by Wildavsky and Dake (Wildavsky and Dake 1990), and that publication suggested that they may be powerful explanatory concepts. However, these results have not been possible to replicate. Correlations between Cultural-Theory scales and perceived risk are typically around 0.05–0.10. This is a very low level of explanatory power. It can be argued that even low correlations may describe relationships of practical value in interventions and experimental designs, but that is a largely irrelevant argument in the present context, where the scales are used in order to test the validity of a theory. Later attempts at operationalizing the concepts of Cultural Theory have not resulted in stronger relationships than the original Wildavsky–Dake scales, see, e.g., Silva et al. (2007). Cultural theory has been found to be useful in work which is more qualitatively oriented, however (West et al. 2010).

Value dimensions in general have been found to be weak explanatory concepts of perceived risk. Other examples than Cultural Theory-scales are the value scales devised by Schwartz (Schwartz 1992), and scales measuring the personality constructs in Jung’s theory of psychological types (Myers et al. 2003). An attempt was made to find relations between Schwartz-type value scales and risk perception of nuclear waste, but no strong relationships were found (Sjöberg 2008b).

A discussion of “distal” explanatory concepts was given by Sjöberg (2003c). The word distal refers to the case when the explanatory factors are semantically divergent from the concept of perceived risk. It is very hard to find strong results with such factors. On the other hand, proximal variables, the opposite of distal variables, should not necessarily be rejected as trivial explanatory concepts. I will return to this issue when I discuss attitude and “affect” later in this chapter.

Interesting, and stronger, results were found with scales measuring “New Age” beliefs (Sjöberg and af Wåhlberg 2002). People who espouse such beliefs tend to be skeptical toward technology. Even stronger relationships have been found with political preferences, measured as party preferences in a multiparty democracy such as Sweden (Sjöberg 2008b). It is likely that values do partly explain party preferences but apparently not values of the kind measured by existing methods to measure value systems, such as the system by Schwartz (Schwartz et al. 2001). Of course, constructs other than values could also be important for political preferences.

The Psychometric Model is another story. It has a high level of explanatory power for *average* ratings of hazards. This is a case of a general principle: averages tend to correlate more strongly than raw data, a phenomenon sometimes referred to as the ecological fallacy (Robinson 1950). However, for *individual* data, the model has only a moderate level of explanatory power, even if it fares better than Cultural Theory. It may, in some cases, explain about 20% of the variance of perceived risk. The most powerful dimension of the Psychometric Model is “dread” (Gardner and Gould 1989).

In spite of occasional successes at the 20% level, the model frequently fails to achieve a better result than Cultural Theory for individual data. Its credibility probably depends of two factors:

1. Data analysis based on averages more or less guarantees high correlations, but it is sometimes not understood that this does not mean that the model explains the risk perception of individuals.
2. The basic dimensions, New risk and Dread, seem to match widespread commonsense notions about lay people's risk perception. There was a particularly powerful demonstration in the basic 1978 paper by Fischhoff et al. which appeared to show that opposition to nuclear power was based on ignorance and emotions.

## Beyond the Psychometric Model

---

As pointed out in the previous section, the traditional models of risk perception explain only a minor part of the variance. In the present section, I briefly review work on other constructs which have led to considerably stronger models. The aim of the work has been to find a model which is as fully explanatory as possible. If a considerable share of the variance of perceived risk is left unexplained, there could be aspects and phenomena which are important but simply not covered at all. The result is likely to be bias in the regression parameters estimated, see e.g., Pedhazur's discussion of specification errors in regression models (Pedhazur 1982, pp. 225–230).

For example, in practical risk perception survey work, which has developed to become a minor "industry," more or less exclusive concern with traditional dimensions of Cultural Theory and the Psychometric Model may have led investigators, and their customers, astray since they have worked with only weak constructs which typically leave about 80% or more of the variance of perceived risk unexplained. It could also be the case that different and more powerful constructs override the effects of the traditional dimensions. For a review of this approach, see Sjöberg (2000b). There are many examples of the low explanatory power of risk perception models, see, e.g., Huang et al. (2010). Eurobarometer work on risk perception of genetically modified food reached only about 30% explained variance (Gaskell et al. 2004); twice as much could be achieved with a less traditional design, see Sjöberg (2008c).

The first successful construct to be mentioned here is that of *risk sensitivity*, see Sjöberg (2004a). People tend to use a risk rating scale in systematically different ways. Some rate all or most hazards very high on the scale, others do the opposite. If all hazard items are treated as part of an attitude scale or as a personality trait, it is always found that they correlate highly and a high alpha value is found. Averaging all items yields a measure of risk sensitivity. It is an interesting topic to inquire into the factors that lead some people to deny or reject all or most hazards as dangerous, while some (usually fewer) do the opposite, see Sjöberg (2006a). The overall risk sensitivity index typically is a very powerful explanatory factor for the risk perceived in any given hazard. Of course, the risk sensitivity index must be adjusted when relating it to any given hazard. The rating of that hazard must be deleted from the index in order to avoid an artifactual relationship.

A second important construct is that of *interfering or tampering with nature*, see Sjöberg (2000c), a third is that of *morality*, see Sjöberg and Winroth (1986). Many people

believe that it is dangerous to interfere with nature, and some also believe that it is morally wrong to do so, possibly because they have religious convictions about nature being created by God. These are powerful factors in perceived risk and attitude toward technologies such as genetically modified food. But they are important also in less obvious cases, such as nuclear technology or nanotechnology (Vandermoere et al. 2010).

Interfering with nature is different from involvement in the environment for two reasons. First, interfering with nature is more widely applicable than involvement with the environment. Second, concern about interference is a quite specific issue and not general, as is the case is with involvement.

What is nature? That is a tricky question. Drottz-Sjöberg and Sjöberg (2009) reported that people will agree that “illness” is part of nature, while the HIV virus is not, in their view. Of course, everything is nature in one sense of the word. To reach a more specific definition, one needs to think hard. It suffices here to point out that people are worried about interference with nature, and that such worries account for a sizable share of the variance of perceived risk.

Nuclear power is at the present becoming more acceptable in many countries. Why? One often mentioned possibility is that the level of perceived risk has decreased, another is connected with benefits. In particular, people may consider that nuclear power is irreplaceable, in the sense that the alternatives are either too costly (solar power, wind) or environmentally unfriendly (fossil fuels). As a more general argument, I show in Sjöberg (2002c) that a technology which was seen as irreplaceable was more readily acceptable, and that this factor contributed strongly beyond perceived risk.

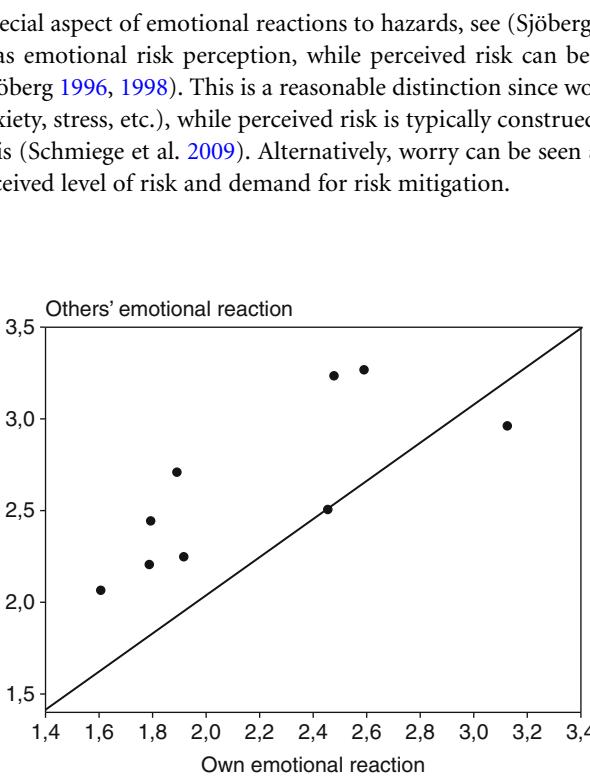
A methodological perspective is brought in by Sjöberg (2003b). In much risk perception work, people were asked to judge “risk” without any further specification. However, it is easy to demonstrate that people make a clear distinction between risk as pertaining to them, personal risk, and risk to others, general risk. The difference has been found to be related to perceived control (Harris and Middleton 1994). People apparently believe that they can control risks, at least some risks, and that others are not willing or able to do so. This is true in particular of lifestyle risks such as smoking or drinking alcohol. When the risk target is not specified, people judge general risks.

Policy implications are different for general and personal risks. Personal risks are important for policy attitudes with regard to environmental and technology risks.

## Affect and Emotion

“Affect” is an ambiguous word (Sjöberg 2006b). It can refer to emotions or values, or both. Emotions are not implicated in studies, where affect is operationalized as value or attitude. It has indeed been found that attitude is a powerful explanatory factor in risk perception; see Sjöberg (1992). The attitude referred to here is that of an object or a construct which is seen as driving the hazard. For example, a nuclear accident is caused by nuclear power. Hence, the perceived risk of a nuclear accident is likely to be found to be explained, to a rather large extent, by the attitude to nuclear power. One could of course question the causal direction here. However, Sjöberg (1992) presents results which support the interpretation that attitude is causally prior to perceived risk.

Emotions are discussed in Sjöberg (2007a). If people are asked to judge their personal emotional reactions to a hazard, it is found that emotions are powerful explanatory concepts of perceived risk. A whole range of emotions (Sherry-Brennan et al. 2010), including fear and anger, are important (Lerner et al. 2003). Positive emotions also enter the picture. Some people react positively to such a concept as nuclear power, finding it interesting or beneficial. The problem with the emotional component in the Psychometric Model (“dread”) is that it is rated for others, not for the respondent him- or herself, and that the model does not include a range of emotions, only one.

It can also be noted that people rate the negative emotions of others as stronger than their own. We tend to believe that others are “emotional” and “irrational,” and that they are more negative toward technology than we are. There is an interesting connection with disaster planning here. “Panic” is expected by many planners, who certainly do not think that they are likely to have such reactions themselves (Wester-Herber, *in press*). Quarantelli showed long ago that this is a myth (Quarantelli 1954), and modern work on risk perception supports his argument.  Figure 25.3 is a scatter plot comparing the strength of anticipated emotional reactions between individuals and others. The x-axis is labeled "Own emotional reaction" and ranges from 1,4 to 3,4. The y-axis is labeled "Others' emotional reaction" and ranges from 1,5 to 3,5. A diagonal line represents the identity line (y=x). Eight data points are plotted, showing that others' emotional reactions are generally stronger than one's own emotional reactions. The data points are approximately at (1.6, 2.1), (1.8, 2.2), (1.8, 2.4), (1.9, 2.7), (2.5, 3.2), (2.5, 3.2), (2.5, 3.3), and (3.2, 2.9).

Worry is a special aspect of emotional reactions to hazards, see (Sjöberg 1998). Worry has been construed as emotional risk perception, while perceived risk can be seen as cognitive (Rundmo and Sjöberg 1996, 1998). This is a reasonable distinction since worry has emotional components (anxiety, stress, etc.), while perceived risk is typically construed as a judgment of how large a risk is (Schmiege et al. 2009). Alternatively, worry can be seen as a driving factor both behind perceived level of risk and demand for risk mitigation.

 Fig. 25.3

Anticipated emotional reactions: own reactions and those of others. Scales refer to strength of anticipated emotional reaction. This figure has been published previously as Fig. 4 in Sjöberg (2007a)

## Trust and Antagonism

Trust has been shown to be an important factor in risk perception (Slovic et al. 1991). It is natural to assume that low trust should be related to a high level of perceived risk. However, the relationship is only moderately strong (Sjöberg and Wester-Herber 2008). The reason appears to be that trust has been too narrowly conceived. Researchers have been mostly interested in social trust, i.e., trust in people and organizations, not in *epistemic* trust or trust in the quality of the knowledge that people and organizations base their risk assessments on.

Drott-Sjöberg found that epistemic trust was frequently mentioned by her respondents in interviews about a nuclear-waste siting project (Drott-Sjöberg 1996, 1998). Sjöberg (2001a) found that epistemic trust was more important than social trust. Effects of social trust on risk perception seemed to be mediated by epistemic trust (Sjöberg 2008a).

Sjöberg (2008a) introduces another concept, namely, *antagonism*; see also Sjöberg and Wester-Herber (2008). Competence is a factor which has been traditionally implicated as important in social trust (Peters et al. 1997). However, perceived competence in an enemy is not reassuring but the opposite. Active enemies have not been traditionally studied in risk perception research; terrorism is a case in point, see Sjöberg (2005). Antagonism is a rather frequently evoked construct among some members of the public, and it contributes to risk perception beyond social and epistemic trust. The relationship between trust and perceived antagonistic interest is illustrated in Fig. 25.4.

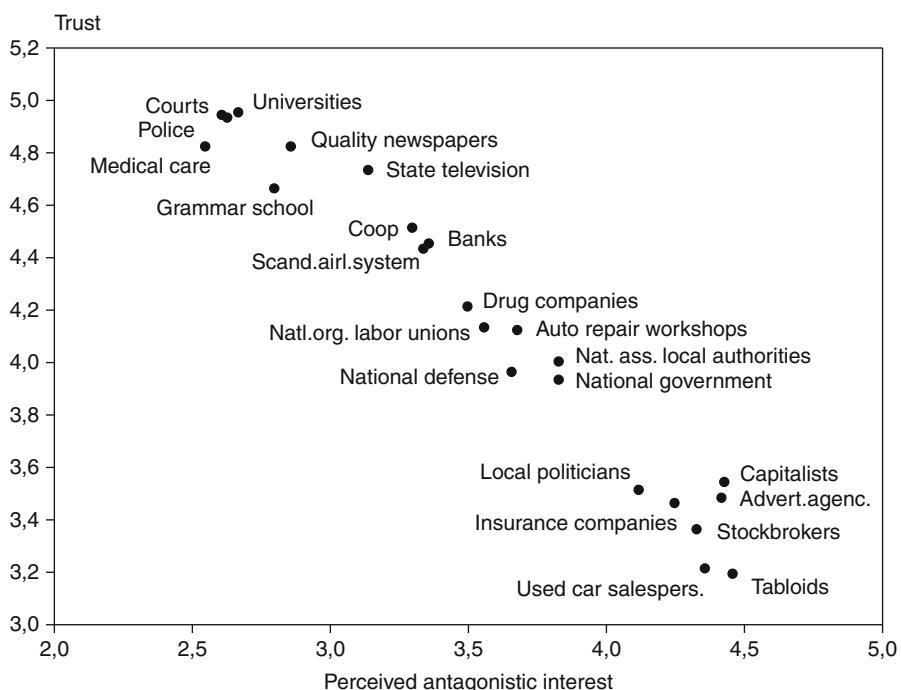


Fig. 25.4

Trust and perceived antagonistic interest of different actors. This figure has been published previously as Fig. 1 in Sjöberg (2008a)

## Risk and Policy

---

One of my earliest papers on risk was an analysis of the conflict between experts and the public about nuclear power (Sjöberg 1980). In the heated debates of the time – and they are not less heated now – both parties were unwilling to listen to each other, and both parties suspected each other of hidden agendas. In the time since then, there have been some notable successes of risk communication, e.g., by the Swedish corporation, SKB, which is working on siting a repository for spent nuclear fuel. This is a notoriously difficult social and political problem in many, perhaps most, countries with nuclear power programs. Only Sweden and Finland have so far been able to proceed to the final stage of the siting process (Sjöberg 2004b, 2008b). The clues to success appear to be:

- Decreased concern over nuclear risks
- Increased emphasis on benefits of nuclear power and worry over climate change
- Information campaigns in the local area
- Open decision making and willingness to let people take part in the process
- An area where many people work in the nuclear industry

Siting issues of this kind are contentious and must be so (Sjöberg 2001b, c). There is no final answer to the size of the “real” risk, and complete unanimity can only be achieved in a totalitarian society, as was the case in Eastern Europe before the fall of the Berlin Wall and the dissolution of the Soviet Union (Sjöberg et al. 2000a). In (Sjöberg 2006a) I tell the story of an impossible “Utopia,” where everybody agrees that risks are large or, more typically, that they are negligible. There are great risks with risk denial. Society tends to regulate risks after the fact, once disasters have happened, rather than plan in order to avoid them. Another paper is about neglected risks and how we can, hopefully, learn from history in order to be more prudent in the future (Sjöberg 2007b). The Precautionary Principle is a dominating policy principle in the European Union. Sjöberg (2009) studied attitudes toward this principle. It was found to be yet another factor contributing to risk perception, probably increasingly important in the future and not only in the EU. Attitudes towards the Precautionary Principle could also be influenced by perceived risk.

The trials and tribulation of risk perception have often been evoked as a way of explaining obvious “irrationalities” of risk policy. One such a case of irrationality has to do with the allocation of economic resources for risk mitigation, see (Sjöberg 1999b). Data from Sweden (Ramsberg and Sjöberg 1997) and other countries (Morral 1986; Tengs et al. 1995) have shown that there are tremendous differences between different hazards and sectors of society with regard to how much society is willing to pay for “saving a life.” In a separate study, we found that people do accept a certain level of variability in allocated resources, but far from what is actually the case (Ramsberg and Sjöberg 1998).

How important is “risk” for policy decisions? For example, do people demand more risk mitigation if the probability of an accident or other type of negative event is higher? It would seem that such must be the case. However, here as so often in risk perception research, things are not what they seem to be (Sjöberg 2003d). People’s judgments of demand for risk mitigation are mostly related to consequences, not to probabilities. And probabilities are almost the same as perceived risks. Two papers (Sjöberg 1999a, 2000a) demonstrate this thesis empirically. Hence, although “risk” is an ambiguous term with components of both probability and consequences, it is interpreted, in natural language discourse, mainly as probability.

This principle can also be demonstrated with a simple-thought experiment: Which risk is largest – to catch a severe cold or to be contaminated with the HIV virus? Most people say “to get a cold.” But which risk should one be most careful to be protected from? Most people say “the HIV virus.” Hence, risk and demand for protection are separated, and demand for protection is governed by the severity of the consequences.

Experts play a central role in risk debates and decision making. There have been studies of experts claiming that they have a qualitatively different approach to risk perception than the public (Slovic et al. 1979). They were said to make risk perception judgments unaffected by the dimensions of the Psychometric Model and to make judgments close to the “true risks.” However, this claim was based on a very small data set and on experts of doubtful competence (Rowe and Wright 2001). I showed that experts had similar risk perception dynamics to those of the public (Sjöberg 2002a). Further data confirming this conclusion is found in Sjöberg and Drottz-Sjöberg (2008a) and Sjöberg (2008c). The latter paper suggests that the original claim of “objective” risk judgments by experts could be due to two factors: they made judgments of “risk” without a specified target, giving general risk judgments rather than personal ones, and their judgments were correlated with “dread” and “new risk” – factors that have a weak explanatory power in accounting for perceived risk (Gardner and Gould 1989). It should be noted that risk due to hazardous technologies are typically related to general risk, while lifestyle risks are related to personal risk (Sjöberg 2003b).

Slovic, Fischhoff, and Lichtenstein (Slovic et al. 1979) found that both experts and lay people made, on the average, reasonably accurate judgments of annual fatalities due to a number of different hazards. Experts’ mean judgments showed a stronger correlation with technical estimates of fatalities, but both sets of data show similar trends in this respect. These results are interesting and reflect what has been called “The wisdom of crowds” by Surowiecki (Surowiecki 2004). However, when the instruction was to rate “risk” rather than the annual number of fatalities the groups diverged strongly, and it was clear that other factors than technical risks entered the judgments made by nonexperts. The effect of such “other factors” can be construed as effects of consequences of accidents rather than probabilities of technical risk estimates.

Risk judgments are typically low when experts judge risks within their area of responsibility, but otherwise they are comparable to risk judgments made by the public. See  Fig. 25.5 which shows large differences between experts and the public, but only for cases where experts have responsibility for risk management.

It is doubtful that the low risk judgments made by experts are mainly based on superior knowledge. It is more likely that they are also the consequence of two factors: self selection and social validation. People seldom enter a profession which they see as dangerous and full of risks to themselves and others (Drottz-Sjöberg and Sjöberg 1991). And once they have been accepted in a professional environment, they are surrounded by others who reinforce their initially low risk perceptions, driving them even lower.

“Stakeholders” constitute a similar case, see Sjöberg (2003a). People, who are actively involved, for or against a contentious risk topic such as a siting of dangerous waste, tend to have views quite different from those of the public at large. For this reason, it is dangerous for decision and policy makers to listen merely to the “stakeholders” who are not representative. “The silent majority” will make its voice heard once it gets a chance to do so, as for example in a local referendum.

Politicians’ risk attitudes and perception have seldom been studied. Sjöberg and Drottz-Sjöberg (2008b) describe an investigation of a large group of Swedish politicians who were

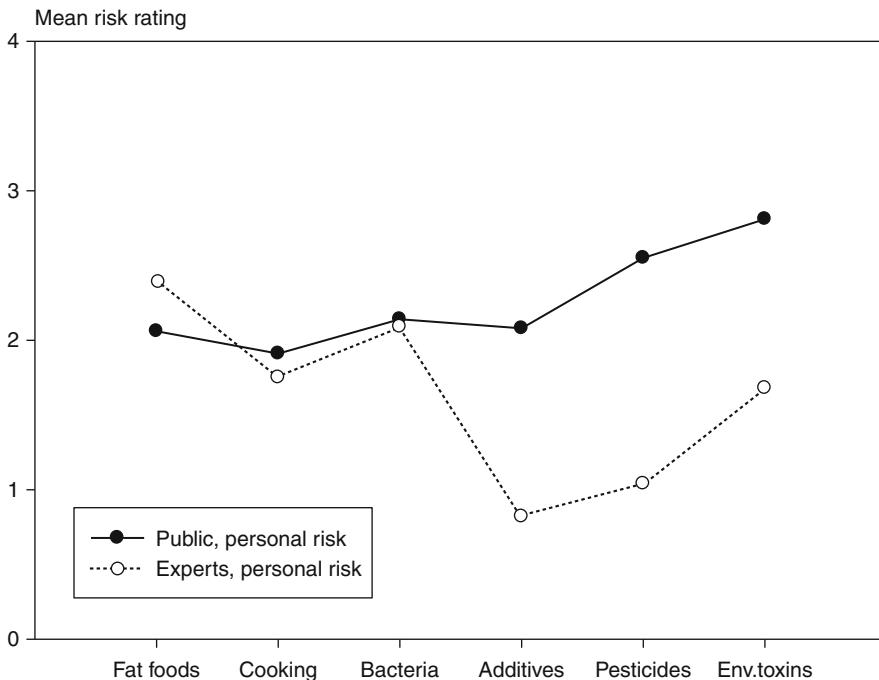


Fig. 25.5

Personal risk ratings, experts, and the public. Y-axis scale refers to size of perceived risk, where 0 = no risk and 5 = a large risk

active in local municipality boards. The major conclusion from this study is that politicians were quite similar to the public in their risk attitudes and perceptions. This was so in spite of the fact that they were older than the average citizen and had a higher level of education. A large proportion of them were men, and men otherwise tend to have lower risk perceptions than women (Davidson and Freudenburg 1996).

## Two Applications

Sjöberg (2008c) and (2005) are reports of risk perception studies of genetically modified food and terrorism, respectively. The first application study shows how the concepts (risk sensitivity, interference with nature, attitude to a hazardous technology, and the distinction between personal and general risks) discussed in the present chapter lead to more powerful models of perceived risk than the traditional ones. It also shows that GM food is a technology which is very little liked by the public in Sweden, similar to what is the case in the EU as a whole and different from the USA. The topic of GM food, and gene technology more generally, is becoming increasingly important. Risk and benefit both play a role, and perhaps benefits are more important than risks for the views of the public. Experts tended to have strongly divergent views as compared to the public. Basic tenets of the experts, such as GM food being in principle no different from food from traditional plant or animal breeding, were simply not believed by the public.

The second application study was about terrorism. There are several new aspects which are brought up by a study of the perceived risk of terrorism (Lee et al. 2010). One is antagonism. Another is general suspiciousness, a personality variable. A third is that of conceptions about the functioning of the terrorists, whether they are rational or not. The study made it very clear that any new major hazard will require its own analysis and new concepts – another important reason why the traditional models of risk perception are insufficient.

## Further Research

---

The differences in risk perception among various groups need more descriptive research and theoretical explanations. Only scientific and technical experts have been investigated to any great extent, while politicians, journalists, and various practitioners such as medical doctors have been little investigated. Yet, all those groups play important roles in societal risk management.

The role of values for risk perception is not well understood. It has been very difficult to find ways of measuring values which show them to be important factors in risk perception. This may be because they in fact are not as important as many people believe them to be, or it may be because our measurement instruments are not good enough, or simply that we do not yet measure the most directly relevant values. Religious values have been little investigated in relation to risk perception.

There has been little work on the life-span development of risk perception from childhood to old age. There are many topics of interest here, such as when gender differences in risk perception tend to appear (Sjöberg and Torell 1993).

Risk perception work originated with the concerns of elites (managers, experts, politicians) over what they saw as irrational worries of the public. Such worries constituted, in the view of many, obstacles to the “rational” exploitation of technologies and natural resources. However, one could turn the question in the opposite direction: why are some people very *unconcerned* about risks? Neglect of risks has sometimes led to disasters. “Whistle-blowers” can make a very useful societal contribution, but they often have to take great personal risks when they do so.

There are many important problems connected with how various groups conceive of each other, such as experts and nonexperts, politicians, and their constituencies.

New technology introduces new risks, and risk perception researchers should be in the forefront of research. Terrorism is one example, nanotechnology another. New concepts are needed. Society faces new and important risks, and the social and psychological factors they bring on to the stage call for evermore research in the fascinating field of risk perception.

## Conclusions

---

A number of principles can be derived from the work reviewed in the present chapter:

1. Personal and general risks differ as to level and, structure; they tend to have different correlates.
2. Models of risk perception need to be extended beyond the Psychometric Model and Cultural Theory. A number of important additional dimensions are mentioned in the chapter (risk sensitivity, interference with nature, attitude to a hazardous technology, and

- the distinction between personal and general risks). The level of explained variance of perceived risk can be enhanced to some 60–70% (Sjöberg 2004a).
3. Trust is important for understanding perceived risk, both social and epistemic. The latter type of trust tends to be the most important one. Perceived antagonistic interest is another important dimension of the general trust concept.
  4. Probabilities and consequences are both important for understanding risk related attitudes, but severity of consequences is more important than probabilities.
  5. Experts tend to have a different risk perspective from that of the public. They rate risks as lower, but only in cases where they have direct responsibility for risk management.
  6. Specific hazards require an analysis in their own right. General risk dimensions are insufficient to catch all the unique aspects brought in by new hazards.

## References

- Breakwell G (2007) The psychology of risk. Cambridge University Press, Cambridge
- Brehmer B (1987) The psychology of risk. In: Singleton WT, Hovden J (eds) Risk and decisions. Wiley, New York, pp 25–39
- Combs B, Slovic P (1979) Newspaper coverage of causes of death. *Journalism Q* 56:837–843,849
- Davidson DJ, Freudenburg WR (1996) Gender and environmental risk concerns – a review and analysis of available research. *Environ Behav* 28:302–339
- Douglas M, Wildavsky A (1982) Risk and culture. University of California Press, Berkeley
- Drottz-Sjöberg B-M (1996) Stämningar i Storuman efter folkomröstningen om ett djupförvar (Moods in Storuman after the repository referendum). Projekt Rapport No. PR D-96-004, SKB, Stockholm
- Drottz-Sjöberg B-M (1998) Stämningar i Malå efter folkomröstningen 1997 (Moods in Malå after the 1997 referendum). Projekt Rapport No. PR D-98-03, SKB, Stockholm
- Drottz-Sjöberg B-M, Sjöberg L (1991) Attitudes and conceptions of adolescents with regard to nuclear power and radioactive wastes. *J Appl Soc Psychol* 21:2007–2035
- Drottz-Sjöberg B-M, Sjöberg L (2009) The perception of risks of technology. In: Grimvall G, Jacobsson D, Thedéen T, Holmgren Å (eds) Risks in technical systems. Springer, New York, pp 255–271
- Fischhoff B, Slovic P, Lichtenstein S, Read S, Combs B (1978) How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. *Policy Sci* 9:127–152
- Gardner GT, Gould LC (1989) Public perceptions of the risk and benefits of technology. *Risk Anal* 9:225–242
- Gaskell G, Allum N, Wagner W, Kronberger N, Torgersen H, Hampel J et al (2004) GM foods and the misperception of risk perception. *Risk Anal* 24:185–194
- Harris P, Middleton W (1994) The illusion of control and optimism about health: on being less at risk but no more in control than others. *Br J Soc Psychol* 33:369–386
- Huang L, Sun K, Ban J, Bi J (2010) Public perception of Blue-Algae bloom risk in Hongze Lake of China. *Environ Manag* 45:1065–1075
- Lee JEC, Lemyre L, Krewski D (2010) A multi-method, multi-hazard approach to explore the uniqueness of terrorism risk perceptions and worry. *J Appl Soc Psychol* 40:241–272
- Lerner JS, Gonzalez RM, Small DA, Fischhoff B (2003) Effects of fear and anger on perceived risks of terrorism: a national field experiment. *Psychol Sci* 14:144–150
- McCombs M, Gilbert S (1986) News influence on our pictures of the world. In: Bryant J, Zillman D (eds) Perspectives on media effects. Erlbaum, Hillsdale, pp 1–15
- Morrall JF III (1986) A review of the record. *Regulation* 10:25–34
- Myers IB, McCaulley MH, Quenk NL, Hammer AL (2003) MBTI manual, 3rd edn. CPP, Palo Alto
- Pedhazur EJ (1982) Multiple regression in behavioral research. Explanation and prediction. Holt, Rinehart and Winston, New York
- Peters RG, Covello VT, McCallum DB (1997) The determinants of trust and credibility in environmental risk communication: an empirical study. *Risk Anal* 17:43–54

- Quarantelli EL (1954) The nature and conditions of panic. *Am J Sociol* 60:265–275
- Ramsberg J, Sjöberg L (1997) The cost-effectiveness of life saving interventions in Sweden. *Risk Anal* 17:467–478
- Ramsberg J, Sjöberg L (1998) The importance of cost and risk characteristics for attitudes towards lifesaving interventions. *Risk Health Saf Environ* 9:271–290
- Robinson WS (1950) Ecological correlations and the behavior of individuals. *Am Sociol Rev* 15:351–357
- Rowe G, Wright G (2001) Differences in expert and lay judgments of risk: myth or reality? *Risk Anal* 21:341–356
- Rundmo T, Sjöberg L (1996) Employee risk perception related to offshore oil platform movements. *Saf Sci* 24:211–227
- Rundmo T, Sjöberg L (1998) Risk perception by offshore oil personnel related to platform movements. *Risk Anal* 18:111–118
- Schmiege SJ, Bryan A, Klein WMP (2009) Distinctions between worry and perceived risk in the context of the theory of planned behavior. *J Appl Soc Psychol* 39:95–119
- Schwartz SH (1992) Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries. In: Zanna MP (ed) *Advances in experimental social psychology*. Academic, San Diego, pp 1–63
- Schwartz SH, Melech G, Lehmann A, Burgess S, Harris M, Owens V (2001) Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *J Cross Cult Psychol* 32:519–542
- Sherry-Brennan F, Devine-Wright H, Devine-Wright P (2010) Public understanding of hydrogen energy: a theoretical approach. *Energy Policy* 38:5311–5319
- Silva C, Jenkins-Smith HC, Barke RP (2007) Reconciling scientists' beliefs about radiation risks and social norms: explaining preferred radiation protection standards. *Risk Anal* 27:758–773
- Sjöberg L (1979) Strength of belief and risk. *Policy Sci* 11:39–57
- Sjöberg L (1980) The risks of risk analysis. *Acta Psychol* 45:301–321
- Sjöberg L (1992) Psychological reactions to a nuclear accident. In: Baarli J (ed) *Conference on the radiological and radiation protection problems in Nordic regions*, Tromsö, 21–22 Nov 1991. Nordic Society for Radiation Protection, Oslo, p 12
- Sjöberg L (1997) Explaining risk perception: an empirical and quantitative evaluation of cultural theory. *Risk Decis Policy* 2:113–130
- Sjöberg L (1998) Worry and risk perception. *Risk Anal* 18:85–93
- Sjöberg L (1999a) Consequences of perceived risk: demand for mitigation. *J Risk Res* 2:129–149
- Sjöberg L (1999b) Life-values and the tyranny of unique decisions. In: Hermerén G, Sahlin N-E (eds) *The value of life*. Royal Academy of Letters, History and Antiquities, Stockholm, pp 73–84
- Sjöberg L (2000a) Consequences matter, "risk" is marginal. *J Risk Res* 3:287–295
- Sjöberg L (2000b) Factors in risk perception. *Risk Anal* 20:1–11
- Sjöberg L (2000c) Perceived risk and tampering with nature. *J Risk Res* 3:353–367
- Sjöberg L (2001a) Limits of knowledge and the limited importance of trust. *Risk Anal* 21:189–198
- Sjöberg L (2001b) Political decisions and public risk perception. *Reliab Eng Syst Saf* 72:115–124
- Sjöberg L (2001c) Whose risk perception should influence decisions? *Reliab Eng Syst Saf* 72:149–152
- Sjöberg L (2002a) The allegedly simple structure of experts' risk perception: an urban legend in risk research. *Sci Technol Hum Value* 27:443–459
- Sjöberg L (2002b) Are received risk perception models alive and well? *Risk Anal* 22:665–670
- Sjöberg L (2002c) Attitudes toward technology and risk: going beyond what is immediately given. *Policy Sci* 35:379–400
- Sjöberg L (2003a) Attitudes and risk perceptions of stakeholders in a nuclear waste siting issue. *Risk Anal* 23:739–749
- Sjöberg L (2003b) The different dynamics of personal and general risk. *Risk Manag Int J* 5:19–34
- Sjöberg L (2003c) Distal factors in risk perception. *J Risk Res* 6:187–211
- Sjöberg L (2003d) Risk perception is not what it seems: the psychometric paradigm revisited. In: Andersson K (ed) *VALDOR conference 2003*. VALDOR, Stockholm, pp 14–29
- Sjöberg L (2004a) Explaining individual risk perception: the case of nuclear waste. *Risk Manag Int J* 6:51–64
- Sjöberg L (2004b) Local acceptance of a high-level nuclear waste repository. *Risk Anal* 24:739–751
- Sjöberg L (2005) The perceived risk of terrorism. *Risk Manag Int J* 7:43–61
- Sjöberg L (2006a) Rational risk perception: Utopia or dystopia? *J Risk Res* 9:683–696
- Sjöberg L (2006b) Will the real meaning of affect please stand up? *J Risk Res* 9:101–108
- Sjöberg L (2007a) Emotions and risk perception. *Risk Manag Int J* 9:222–237
- Sjöberg L (2007b) Försummade risker. (Neglected risks). In: Derefeldt G, Sjöstedt G (eds) *SDSS Årsbok 2007. Strukturerad osäkerhet, ostrukturera säkerhet i en globaliserad värld*. Utrikespolitiska institutet, Stockholm, pp 39–51

- Sjöberg L (2008a) Antagonism, trust and perceived risk. *Risk Manag Int J* 10:32–55
- Sjöberg L (2008b) Attityd till slutförvar av använt kärnbränsle: Struktur och orsaker (Attitudes toward the final repository for spent nuclear power: Structure and causes). Research Report No. R-08-119, SKB, Stockholm. Svensk Kärnbränslehantering AB
- Sjöberg L (2008c) Genetically modified food in the eyes of the public and experts. *Risk Manag Int J* 10:168–193
- Sjöberg L (2009) Precautionary attitudes and the acceptance of a local nuclear waste repository. *Saf Sci* 47:542–546
- Sjöberg L, af Wählberg A (2002) Risk perception and new age beliefs. *Risk Anal* 22:751–764
- Sjöberg L, Drott-Sjöberg B-M (2008a) Attitudes towards nuclear waste and siting policy: experts and the public. In: Lattefer AP (ed) Nuclear waste research: siting, technology and treatment. Nova Publishers, New York, pp 47–74
- Sjöberg L, Drott-Sjöberg B-M (2008b) Risk perception by politicians and the public. *Energy Environ* 19:455–483
- Sjöberg L, Engelberg E (2010) Risk perception and movies: a study of availability as a factor in risk perception. *Risk Anal* 30:95–106
- Sjöberg L, Torell G (1993) The development of risk acceptance and moral valuation. *Scand J Psychol* 34: 223–236
- Sjöberg L, Wester-Herber M (2008) Too much trust in (social) trust? The importance of epistemic concerns and perceived antagonism. *Int J Glob Environ Issue* 30:30–44
- Sjöberg L, Winroth E (1986) Risk, moral value of actions, and mood. *Scand J Psychol* 27:191–208
- Sjöberg L, Kolarova D, Rucai A-A, Bernström M-L (2000a) Risk perception in Bulgaria and Romania. In: Renn O, Rohrmann B (eds) Cross-cultural risk perception. A survey of empirical studies. Kluwer, Dordrecht, pp 145–184
- Sjöberg L, Truedsson J, Frewer LJ, Prades A (2000b) Through a glass darkly: experts' and the public's mutual risk perception. In: Cottam MP, Harvey DW, Pape RP, Tait J (eds) Foresight and precaution, vol 1. A. A. Balkema, Rotterdam, pp 1157–1162
- Slovic P, Fischhoff B, Lichtenstein S (1979) Rating the risks. *Environment* 21(14–20):36–39
- Slovic P, Flynn JH, Layman M (1991) Perceived risk, trust, and the politics of nuclear waste. *Science* 254:1603–1607
- Surowiecki J (2004) The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations. Doubleday, New York
- Tengs OT, Adams ME, Pliskin JS, Safran DG, Siegel JE, Weinstein MC et al (1995) Five-hundred life saving interventions and their cost effectiveness. *Risk Anal* 15:369–390
- Tversky A, Kahneman D (1973) Availability: a heuristic for judging frequency and probability. *Cogn Psychol* 4:207–232
- Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. *Science* 185: 1124–1131
- Vandermoere F, Blanchetanche S, Bieberstein A, Marette S, Roosen J (2010) The morality of attitudes toward nanotechnology: about God, techno-scientific progress, and interfering with nature. *J Nanopart Res* 12:373–381
- West J, Bailey I, Winter M (2010) Renewable energy policy and public perceptions of renewable energy: a cultural theory approach. *Energy Policy* 38: 5739–5748
- Wester-Herber M, Fight, flight or freeze: assumed reactions of the public during a crisis. *J Contingen Crisis Manage* (in press)
- Wildavsky A, Dake K (1990) Theories of risk perception: who fears what and why? *Daedalus* 119: 41–60



# 26 The Role of Feelings in Perceived Risk

*Melissa L. Finucane*

East-West Center, Honolulu, HI, USA

<b><i>Introduction</i></b> .....	<b>678</b>
<b><i>Conceptualizations</i></b> .....	<b>678</b>
Dual-Process Theories: Recognizing Reliance on Feelings .....	678
Functional Frameworks: Identifying the Roles of Feelings .....	680
Clarifying the Relationship Between Feelings-Based and Cognitive Processes .....	681
<b><i>Empirical Support</i></b> .....	<b>682</b>
Integral Feelings as a Proxy for Value .....	682
Psychophysical Numbing .....	684
Nonintuitive Consequences .....	685
Misattribution of Incidental Feelings .....	686
<b><i>Generalizations</i></b> .....	<b>687</b>
<b><i>Further Research</i></b> .....	<b>688</b>

**Abstract:** This chapter provides an overview of key conceptualizations of and evidence for the role of feelings in perceived risk. Influence from feelings in judgment and decision making was first recognized nearly three decades ago. More recent work has developed models that generalize the mechanisms by which feelings operate. Feelings may play multiple roles in judgment and decision processes, including providing information, enabling rapid information processing, directing attention to relevant aspects of the problem, facilitating abstract thought and communication, and helping people to determine social meaning and to act morally. Feelings may be anticipated or experienced immediately and either integral (attached) to mental representations of the decision problem or incidental (unrelated), arising from moods or metacognitive processes. A rich repertoire of psychological concepts related to risk, such as appraisal and memory, can be used to help explain the mechanisms by which affect and analysis might combine in judgment and decision making. Phenomena such as psychophysical numbing, probability neglect, scope insensitivity, and the misattribution of incidental affect all provide empirical support, albeit fragmented, for the important influence of feelings. Future research needs to utilize multiple dependent variables and methodological approaches to provide convergent evidence for and development of more sophisticated descriptive and predictive models. An additional direction for future research is to develop tools that help risk communicators and risk managers to address complex, multidimensional risk problems.

## Introduction

---

The study of the role of feelings in risk judgments began with a focus on regret and disappointment theories within an economic framework (Bell 1982; Loomes and Sugden 1982) and experimental manipulations of mood (Johnson and Tversky 1983; Isen and Geva 1987). Nearly three decades later, researchers have amassed considerable evidence recognizing the importance of feelings in shaping risk perceptions. Numerous approaches have been used to capture and explicate feelings-based processes in a wide variety of domains. Research has moved on from establishing that feelings play a role, to developing models that generalize the mechanisms by which risk perceptions are influenced (Pham 2007; Slovic 2010). This chapter provides an overview of key conceptualizations of and evidence for the role of feelings in risk judgments. An intentionally wide-ranging use of the term “feelings” is employed to include studies of affect, emotion, and mood, reflecting the diverse theories and methods that comprise this field of research.

## Conceptualizations

---

### Dual-Process Theories: Recognizing Reliance on Feelings

---

Neoclassical economics asserts that individuals, over time and in aggregate, process risk information only in a way that maximizes expected utility (von Neumann and Morgenstern 1947). From this perspective, judgments are based on a utilitarian balancing of risks and benefits and feelings are only a byproduct of the cognitive process. That is, emotions such as fear, dread, anger, hope, or relief are experienced *after* the risk-benefit calculation is complete.

More recently, dual-process theories have conceptualized perceptions of and responses to risk as typically reflecting two, interacting, information-processing systems (Damasio 1994; Epstein 1994; Sloman 1996; Kahneman 2003; Bechara and Damasio 2005). The “analytic” system reflects the slow, deliberative analysis we apply to assessing risk and making decisions about how to manage hazards. The “experiential system” reflects fast, intuitive, affective reactions to danger. “Affective reactions” refer to a person’s positive or negative feelings about specific objects, ideas, images, or other target stimuli. Feelings may also reflect emotions (intense, short-lived states of arousal accompanied by expressive behaviors, specific action tendencies, and conscious experiences, usually with a specific cause,Forgas 1992) and moods (feelings with low intensity, lasting a few minutes or several weeks, often without specific cause, Isen 1997). From the dual-process perspective, feelings that arise from or amidst the experiential mode of thinking are influential *during* judgment and decision-making processes (Schwarz and Clore 1988).

Reliance on feelings in the process of evaluating risk has been termed “the affect heuristic” (Finucane et al. 2000a). Feelings provide potentially useful inputs to judgments and decisions, especially when knowledge about the events being considered is not easily remembered or expressed (Damasio 1994). Many theorists have also given feelings a direct and primary role in motivating and regulating behavior (Mowrer 1960; Zajonc 1980; Damasio 1994; Isen 1997; Kahneman 2003; Pham 2007). Positive feelings act like a beacon of incentive, motivating people to act to reproduce those feelings, whereas negative feelings motivate actions to avoid those feelings.

Since recognizing the importance of feelings, scholars have attempted to clarify the nature and timing of their influence on risk perceptions. Distinguishing the impact of specific emotional states is of concern because the desirability of the impact may be a function of the intensity, valence, and appraisal content of the emotion. For instance, Lerner and Keltner (2001) have shown that fearful people express pessimistic risk estimates and risk-averse choices, whereas angry people express optimistic risk estimates and risk-seeking choices. Similarly, the importance of anticipated regret and disappointment has been demonstrated by Zeelenberg et al. (2000) and Connolly and Butler (2006). The timing of feelings is also critical. In analyses of the time course of decisions, Loewenstein and Lerner (2003) distinguish between anticipated emotions (beliefs about one’s future emotional states that might ensue after particular outcomes) and immediate emotions (experienced when making a decision, thereby exerting an influence on the choice process). (For similar distinctions see Kahneman 2000.) Loewenstein and Lerner further identified two types of immediate emotions, namely, integral emotions (caused by the decision problem itself, such as feelings about a target stimulus or available options) and incidental emotions (caused by factors unrelated to the decision problem at hand, such as mood or cognitive fluency); see also Bodenhausen (1993) and Pham (2007). Empirical demonstrations of the influence of integral and incidental feelings on a wide variety of judgments and decisions are reviewed below.

In sum, early models of judgment and decision making emphasized cognitive aspects of information processing and viewed feelings only as a byproduct of the cognitive process. More recent models, however, give feelings a direct and primary role in motivating and regulating behavior in response to risk. Feelings may be anticipated or immediate and either integral to the decision problem or incidental, arising from moods or metacognitive processes. Identifying the role of feelings and how they interact with cognitive processes is the current focus of scientific inquiry for many researchers.

## Functional Frameworks: Identifying the Roles of Feelings

---

In recent work, Peters (2006) proposed a framework to capture four roles that feelings play in judgment and decision processes. The first role is to provide information about the target being evaluated. Based on prior experiences relevant to choice options (integral affect) or the result of less relevant and ephemeral states (incidental affect), feelings act as information to guide the judgment or decision process (Slovic et al. 2002). The second role is as a spotlight. The extent or type of feelings (e.g., weak vs. strong or anger vs. fear) focuses the decision maker's attention on certain kinds of information, making it more accessible for further processing. Third, feelings may operate as a motivator of information processing and behavior, influencing approach-avoidance tendencies (Frijda et al. 1989; Zeelenberg et al. 2008). Incidental mood states may also motivate people to act in a way that maintains a positive mood (Isen 2000). A fourth role is to serve as a common currency in judgments and decisions, allowing people to compare disparate events and complex arguments on a common underlying dimension (Cabanac 1992). Integrating good and bad feelings is easier than trying to integrate multiple incommensurate values and disparate logical reasons. A similar functional framework has been proposed by Pfister and Böhm (2008), who emphasize the role of feelings in providing information, directing attention to relevant aspects of the problem, and enabling rapid information processing.

An additional function of feelings, according to Pfister and Böhm (2008), is to generate commitment to implementing decisions, thus helping people to act morally, even against their short-term self-interest. Roeser (2006; 2009; 2010) also highlights the importance of emotions in providing moral knowledge about risks and that emotions are needed to correct immoral emotions. Kahan (2008) describes emotion as providing "a perceptive faculty uniquely suited to discerning what stance toward risk best coheres with a person's values." In his cultural evaluator theory, Kahan regards emotion as entering into risk judgments as a way of helping people to evaluate the social meaning of a particular activity against a background of cultural norms and to express the values that define their identities. When people draw on their feelings to judge risk, they form an attitude about what it would mean for their cultural worldviews for society to agree that the risk is dangerous and worthy of regulation. Kahan distinguishes the role of feelings not as a heuristic but as unique in enabling a person to identify a stance that is "expressly rational" for someone with commitment to particular worldviews. Consistent with his theory, Kahan et al. (2007) found that the impact of affect relative to other influences (such as gender, race, or ideology) was significantly larger among people who knew a modest or substantial amount about nanotechnology. This contrasts with the heuristic perspective in which affect is expected to play a larger role when someone lacks sufficient information to form a coherent judgment.

Finally, feelings may help to facilitate abstract thought and communication (Finucane and Houpis 2006). Feelings help people to think abstractly because they link abstract concepts (e.g., good, bad) to the physical or sensory world. Without such links, judgments are slower and less accurate. One subtle demonstration of the link between affect and analytic thought is research showing that positive words are evaluated faster and more accurately when presented in white font, whereas negative words are evaluated faster and more accurately when presented in black font, despite the brightness manipulation being orthogonal to the valence of the words (Meier and Robinson 2004). Similarly, Meier and Robinson (2004) showed that people assign "goodness" to objects high in visual space and "badness" to objects low in visual space. Linking abstract concepts to physical or sensory experiences helps the analytic system to interpret the meaning of stimuli so that they can be incorporated in cognitive calculus.

In sum, several roles of feelings in judgments and decisions about risk have been identified. Additional roles may be articulated as diverse disciplines apply their perspectives. Which role dominates in any particular judgment or decision is likely to be a function of multiple factors (e.g., task demands, time pressure, preferred decision style, social norms).

## Clarifying the Relationship Between Feelings-Based and Cognitive Processes

---

Despite recognition that feelings-based and cognitive processes represent interdependent systems in decision making (Damasio 1994; Epstein 1994; Sloman 1996; Kahneman 2003; Bechara and Damasio 2005), theory and research to date have struggled to convey the exact nature of the relationship. The cognitive origins of behavioral decision theory may have encouraged people to assume that the domain of feelings is qualitatively different and functionally separate from the domain of cognition. Such distinction is reflected in the dichotomies often portrayed in this field, such as irrational emotions disturbing rational cognitions, intuitive feelings dominating deliberate thinking, and hot affect overwhelming cold logic (Pfister and Böhm 2008). However, overlapping commonalities in the systems have been noted also. For instance, the processing of experiences may be involved in both affective and analytic approaches. Johnson and Lakoff (2002; Lakoff and Johnson 1999) point out that even our most abstract thinking (mathematics, for example) is based on our “embodied” experiences. They describe how the locus of experience, meaning, and thought is the ongoing series of physical interactions with our changing environment. Our embodied acts and experiences are an important part of our conceptual system and in making sense of what we experience.

Clarifying the mechanisms by which feelings and cognitions are related and integrated in human judgment and decision making is a critical next step in understanding perceived risk. Finucane and Houpis (2006) recommend expanding and linking the risk-as-analysis and risk-as-feelings approaches by adopting a “risk-as-value” model. This model emphasizes that responses to risk result from a combination of analysis and affect that motivates individuals and groups to achieve a particular way of life. Derived from dual-process theories, the risk-as-value model implies that differences in perceived risk may arise from differences in the analytic or affective evaluation of a risk or the way these evaluations are combined. As research moves from simply describing variance to predicting it, having multiple potential loci for such variation with different substantive interpretations will be useful. The risk-as-value model does not posit a specific rule for combining affective and analytic evaluations, although traditional information integration rules (adding, averaging, multiplying) may be applicable in some contexts. When the implications of both affective and analytic evaluations are congruent, the processes may be more likely to combine additively. However, incongruence may result in greater emphasis on analytic or affective processing, depending on an array of task, decision-maker, or context variables (e.g., analysis may be increased if it is viewed as more reliable, but may be attenuated under time pressure).

The relationship between affective and analytic processes may be more fully explained by drawing on the rich repertoire of empirically tested concepts related to the psychology of risk, such as appraisal and memory. Lerner and Keltner’s (2000) appraisal-tendency theory suggests that emotions arise from but also elicit specific cognitive appraisals. For instance, fear

arises from and evokes appraisals of uncertainty and situational control, whereas anger is associated with appraisals of certainty and individual control (Lerner and Keltner 2001). Lerner et al. (2003) showed that anger evokes more optimistic beliefs about risks such as terrorism, whereas fear evokes greater pessimism about risks. Weber and Johnson's (2006) preferences-as-memory framework highlights how risk judgments are made by retrieving relevant (cognitive and affective) knowledge from memory. Framing normatively equivalent information positively or negatively (e.g., 90% lives saved vs. 10% lives lost) influences preferences because the different descriptions prime different representations in memory (predominantly positively or negatively valenced). Also drawing on modern concepts of memory representation, retrieval, and processing, Reyna and colleagues (Reyna and Brainerd 1995; Reyna et al. 2003) have proposed a dual-process model called fuzzy-trace theory (FTT). FTT posits that people form two kinds of mental representations. The first, verbatim representations, are detailed and quantitative. The second, gist representations, provide only a fuzzy trace of experience in memory. People tend to rely primarily on gist, which captures the meaning of experience, including the emotional meaning. FTT differs from other dual-process models by placing intuition at the highest level of development, viewing fuzzy intuitive processes as more advanced than precise analytic processes (Reyna 2004).

In sum, the mechanisms by which feelings and cognitions are combined in judgment and decision making need to be clarified. Studies from a wide range of disciplines, including cognitive and social psychology, emotion and motivation, economics, decision research, and neuroscience, need to be integrated to develop models that explicitly specify possible causal constructs or variables that influence reactions to risk, allow for individual and group differences in these variables or in the relationships between them, and generalize across risk domains and contexts. Such model-based research can broaden our understanding of risk perceptions specifically and of basic psychological phenomena more generally.

## **Empirical Support**

---

This section briefly reviews empirical support for the role of feelings in risk judgments and decisions. Although the empirical literature seems fragmented and sometimes inconsistent, evidence for the influence of emotion, affect, and mood is compelling.

### **Integral Feelings as a Proxy for Value**

---

Early evidence of the role of feelings in risk perceptions came from studies showing that “dread” was the major driver of public acceptance of risk in a wide range of contexts, including environmental hazards such as pesticides, coal burning (pollution), and radiation exposure from nuclear power plants (Fischhoff et al. 1978). This observation led to many studies looking at how risk judgments are influenced by feelings that are integral (attached) to mental representations of hazardous activities, technologies, or events (Loewenstein et al. 2001; Slovic et al. 2002). In the first paper published on the affect heuristic, Finucane et al. (2000a) demonstrated that providing information about benefit (e.g., of nuclear power) changed perceptions of risk and vice versa. They also showed that whereas risk and benefit (e.g., of natural gas, chemical fertilizers) tend to be positively correlated across hazards in the world,

they are negatively correlated in people's judgments. Moreover, this inverse relationship between perceived risks and benefits increased greatly under time pressure, a situation in which opportunity for analytic deliberation was reduced. Although subconscious cognitive processes cannot be ruled out entirely, these results support the notion that in the process of judging risk, people may rely on feelings as a source of information about whether or not they are at risk and how they should respond.

Underpinning processes such as the affect heuristic are images, to which positive or negative feelings become attached through learning and experience. Images include perceptual representations (pictures, sounds, smells) and symbolic representations (words, numbers, symbols) (Damasio 1999). In an influential series of studies using the Iowa Gambling Task, Damasio, Bechara, and colleagues (Bechara et al. 1994; Damasio 1994; Bechara et al. 1997) proposed that in normal individuals, emotional responses evoked by objects are stored with memory representations (images) as somatic markers of these objects' value (for challenges to the original interpretation, however, see Maia and McClelland 2004; Fellows and Farah 2005). Other research suggests that more vivid, emotionally gripping images of harm are more salient than emotionally sterile images, making those risks more likely to be noticed, recalled, and responded to (Hendrickx et al. 1989; Sunstein 2007). One explanation for this vividness effect may be that initial affective responses to an object seem to trigger a confirmatory search for information that supports the initial feelings (Pham et al. 2001; Yeung and Wyer 2004), possibly increasing the subjective coherence of judgments based on affect (Pham 2004). Another explanation may relate to the inherently strong drive properties of integral feelings, which motivate behavior and redirect action if necessary (Frijda 1988).

A simple method for studying the relationship between affect, imagery, and perceived risk is called affective image analysis, a structured form of word association and content analysis (Slovic et al. 1991; Benthin et al. 1995; Finucane et al. 2000b; Jenkins-Smith 2001; Satterfield et al. 2001). This method allows researchers to examine the distribution of different (sometimes conflicting) meanings of risk across people and to identify and explain those images that carry a strongly positive or negative emotional charge. For instance, Finucane et al. (2000) asked study participants to free associate to the phrase "blood transfusions." Associations included "HIV/AIDS," "hemophilia," "gift giving" and "life saving." Participants were then asked to rate each of their associations on a scale from bad (-3) to good (+3); these ratings were correlated with a number of other measures, such as acceptability of having a transfusion and sensitivity to stigmatization in other risk settings. Affective image analysis was also used in a US national survey by Leiserowitz (2006), who found that holistic negative affect and image affect were significant predictors of global warming risk perceptions, explaining 32% of the variance. Holistic negative affect was also predictive of support for national policies to address global warming, but less predictive than worldviews and values. A content analysis of affective imagery associated with "global warming" revealed that the phrase evoked negative connotations for almost all respondents, but that the most dominant images referred to impacts that were psychological or geographically distant, generic increases in temperature, or a different environmental problem.

In sum, integral affective responses are feelings elicited by real, perceived, or imagined images of the object of judgment or decision. These feelings are predictive of a variety of behavioral responses to risk. Evaluation and choice processes are more likely to be influenced by vivid, emotionally gripping images than pallid representations, possibly because strong feelings trigger a confirmatory information search or strong drive states.

## Psychophysical Numbing

---

Considerable evidence suggests that affective responses follow the same psychophysical function that characterizes our sensitivity to a range of perceptual stimuli (e.g., brightness, loudness). In short, people's ability to detect changes in a physical stimulus decreases as the magnitude of the stimulus increases. Known as Weber's law, the just-noticeable change in a stimulus is a function of a fixed percentage of the stimulus. That is, to notice a change, only a small amount needs to be added to a small stimulus, but a large amount needs to be added to a large stimulus (Stevens 1975). Our cognitive and perceptual systems are designed to detect small rather than large changes in our environment. Fetherstonhaugh et al. (1997) demonstrated this same phenomenon of psychophysical numbing (i.e., diminished sensitivity) in the realm of feelings by evaluating people's willingness to fund alternative life-saving medical treatments. Study participants were asked to indicate the number of lives a hypothetical medical research institute would have to save to merit a \$10 million grant. Nearly two-thirds of participants raised their minimum benefit requirements to warrant funding when the at-risk population was larger. A median value of 9,000 lives needed to be saved when 15,000 were at risk, compared with a median of 100,000 lives when 290,000 were at risk. In other words, 9,000 in the smaller population seemed more valuable than saving ten times as many lives in the larger population. Psychophysical numbing or proportional reasoning effects have been demonstrated also in other studies (Baron 1997; Friedrich et al. 1999).

In striving to explain when feelings are most influential in judgments about saving human lives, several researchers have explored the "identifiable victim effect" (Jenni and Loewenstein 1997; Kogut and Ritov 2005; Small and Loewenstein 2005). For instance, Small et al. (2007) asked participants to indicate how much they would donate to a charity after being shown either statistical information about the problems of starvation in Africa ("statistical victims") or a photograph of a little girl in Africa and a brief description of the starvation challenges she faces ("identifiable victim"). Results showed that the mean donation (\$2.83) for the identifiable victim was more than twice the mean donation (\$1.17) for the statistical victim, as might be expected given the affectively engaging nature of the photograph of the identifiable victim. Most interestingly, however, when participants were shown both statistical and identifiable information simultaneously, the mean donation was \$1.43. When jointly evaluating statistics and an individual victim, the reason for donating seems to become less compelling, possibly because the statistics diminish reliance on affective reactions during decision making. Small et al. also measured feelings of sympathy toward the cause (the identified or statistical victims). The correlation between these feelings and donations was strongest when people faced the identifiable victim.

In a follow-up study by Small et al. (2007), participants were either primed to feel ("Describe your feelings when you hear the word 'baby'") or to deliberate ("If an object travels at five feet per minute, how many feet will it travel in 360 s?"). Relative to the feelings prime, priming deliberative thinking reduced donations to the identifiable victim. There was no discernable difference of the two primes on donations to statistical victims, as would be expected because of the difficulty in generating feelings for such victims. Similarly, Hsee and Rottenstreich (2004) demonstrated that priming analytic evaluation led to more scope sensitivity and affective evaluation led to more scope insensitivity when participants were asked how much they would be willing to donate to help save endangered pandas. In their study, the number of pandas was represented in an affect-poor manner (i.e., as large dots) or an affect-rich manner (i.e., with a cute picture). The dots were related to a fair degree of scope sensitivity

(mean donations were greater for four pandas than one), whereas pictures were related to scope insensitivity (mean donations for four versus one panda were almost identical). This scope insensitivity violates logical rationality, suggesting that inherent biases in the affective system can lead to faulty judgments and decisions.

In sum, the affective system seems designed to be most sensitive to small changes at the cost of making us less able to respond appropriately to larger changes further away from zero. Consequently, we may fail to respond logically to humanitarian and environmental crises.

## Nonintuitive Consequences

---

Integral affect may lead decision makers astray in several other ways. One example is the phenomenon of “probability neglect” – the failure of people to adjust their decisions about the acceptability of risks to changes in information about their probability. Loewenstein et al. (2001) observed that responses to uncertain situations appear to have an all-or-none characteristic, sensitive to the possibility of strong negative or positive consequences and insensitive to their probability. That is, strong feelings tend to focus people on outcomes rather than probabilities. Rottenstreich and Hsee (2001) demonstrated that while people were willing to pay more to avoid a high than a low probability of losing \$20, they were not willing to pay more to avoid a high than a low probability of receiving an electric shock (a prospect rich in negative affect). Another example comes from Denes-Raj and Epstein’s (1994) jellybeans experiment. When given a chance to draw a winning red bean either from a small bowl containing a single red bean and nine white beans (10% chance of winning) or from a larger bowl containing nine red beans and 91 white beans (9% chance of winning), people tend to choose to draw from the larger bowl, even though the probability of winning is greater with the small bowl. The more abstract notion of probability (the distribution of beans in a random draw process) is less influential than the affective response people have to the concrete representations of objects (seeing multiple red beans). One interpretation of these results is that integral affect provides a largely categorical approach to assessing value. That is, objects are categorized in terms of their significance for well-being, regardless of their probability or magnitude.

The emerging field of neuroeconomics provides convergent evidence for the nonintuitive consequences of integral affect (Trepel et al. 2005). Using methods such as functional magnetic resonance imaging, researchers have examined brain activity in areas known to process affective information. For instance, examining the neurobiological substrates of dread, Berns et al. (2006) showed that when people are confronted with the prospect of an impending electric shock, regions of the pain matrix (a cluster of brain regions activated during a pain experience) are activated. This finding suggests that people not only dislike experiencing unpleasant outcomes, they also dislike waiting for them. Contrary to tenets of economic theory, people seem to derive pain (and pleasure) directly from information, rather than from any material outcome that the information might lead to. Anticipating future outcomes in this way can have a major impact on intertemporal choices (decisions that involve costs and benefits that extend over time). While an economic account of intertemporal choice predicts that people generally want to expedite pleasant outcomes and delay unpleasant ones (Loewenstein 1987), an affective account predicts that people may prefer to defer pleasant outcomes when waiting is pleasant or to expedite unpleasant outcomes when waiting is frustrating or produces dread.

Another nonintuitive feature of feelings-based judgments is that they tend to be more relativistic or reference-dependent than are reason-based judgments. That is, affective responses are often not based on the object or outcome in isolation, but in relation to other objects or outcomes (Mellers 2000). Winning \$10 in a gamble will elicit greater pleasure if the alternative outcome is losing \$5 rather than only \$1. This finding is also consistent with work on the evaluability principle. Hsee (1996) asked people to assume they were music majors looking for a used music dictionary. Participants were shown two dictionaries and asked how much they would be willing to pay for each. Dictionary A had 10,000 entries and was like new, whereas dictionary B had 20,000 entries but also had a torn cover. In a joint-evaluation condition, willingness to pay was higher for B, presumably because of its greater number of entries. However, when one group of participants evaluated only A and another group evaluated only B, the mean willingness to pay was much higher for A, presumably because without a direct comparison, the number of entries was hard to evaluate whereas the defects attribute was easy to translate into a precise good/bad response. Wilson and Arvai (2006) have extended this work to show that in some contexts, enhanced evaluability may not be sufficient to deflect attention away from the affective impressions of the choice pair and toward other decision-relevant risk information, a behavior they call affect-based value neglect.

In sum, strong feelings can lead people to ignore probabilities and magnitudes, possibly because in some situations integral affect can provide only a categorical and reference-dependent approach to valuation. Risk theory and practice will benefit from further explorations of the conditions under which feelings influence attention to and use of different types of information.

## Misattribution of Incidental Feelings

---

In addition to studies focusing on integral feelings, a large number of studies have shown that affective states unrelated to the judgment target (incidental feelings) may influence judgments and decisions (Schwarz and Clore 1983; Isen 1997). An early study by Johnson and Tversky (1983) demonstrated that experimental manipulation of mood (induced by a brief newspaper report on a tragic event such as a tornado or flood) produced a pervasive increase in frequency estimates for many undesirable events, regardless of the similarity between the report and the estimated risk. More recently, Västfjäll et al. (2008) showed that eliciting negative affect in people by asking them to think about a recent major natural disaster (the 2004 tsunami) influenced judgments when the affect was considered relevant (e.g., the perceived risk of traveling to areas affected by the disaster), but also when it was not relevant (e.g., developing gum problems).

In a classic study, Schwarz and Clore (1983) demonstrated that people reported higher levels of life satisfaction when they were in a good mood as the result of being surveyed on a sunny day than people who were in a bad mood as a result of being surveyed on a rainy day. People incorrectly attributed their incidental moods as a reflection of how they felt about their personal lives. In general, the misattribution of incidental feelings to attentional objects tends to distort beliefs in an assimilative fashion. However, research suggests that the influence of incidental affect is not stable nor unchangeable. Rather, it is a constructive process in which the decision maker needs to determine whether their feelings are a reliable and relevant source of information (Pham et al. 2001; Clore and Huntsinger 2007). For instance, Schwartz and Clore

were able to reduce the influence of mood on participants' judgments of well-being with a simple reminder about the cause of their moods (e.g., sunny vs. cloudy weather), presumably triggering people to question the diagnostic value of the affective reaction for the judgment. Importantly, the manipulation changed the diagnostic value of the affective reaction, not the affective reaction itself (Schwarz 2004). Västfjäll et al. (2008) also demonstrated that manipulating the ease with which examples of disasters come to mind can influence risk estimates. Asking participants who had been reminded of the 2004 tsunami to list few (vs. many) natural disasters led to more pessimistic outlooks (measured via an index averaging judgments of the likelihood of positive and negative events), presumably because listing many natural disasters rendered incidental affect relatively less diagnostic for judgments.

Incidental affective states have been shown also to influence the nature of information processing most likely to occur. Negative mood states generally promote a more analytic form of information processing, whereas positive moods generally promote a less systematic, explorative form of processing. From an evolutionary perspective, negative moods may highlight a discrepancy between a current and desired state, signaling a need to analyze the environment carefully (Higgins 1987). Positive moods, on the other hand, may encourage variety seeking in order to build future resources (Fredrickson 1998). Empirical findings are not entirely consistent, however. Both positive and negative moods have been related to increased and decreased systematic processing (Isen and Geva 1987; Mackie and Worth 1989; Schwarz 1990; Baron et al. 1992; Gleicher and Petty 1992; Wegener and Petty 1994; Isen 1997).

In sum, incidental feelings may influence risk judgments and decisions. The diagnostic value of the feelings depends on the context. Fortunately, people can be primed to examine the diagnostic value of their feelings. Incidental feelings may also influence the extent to which individuals engage in systematic processing, although the exact nature of this relationship remains unclear.

## Generalizations

---

Several generalizations can be made about the role of feelings in risk judgments. First, feelings in the form of emotions, affect, or mood can have a large impact on how risk information is processed and responded to. The multiple ways in which feelings influence risk judgments and decisions likely relate to several functions of feelings: providing information, focusing attention, motivating behavior, enabling rapid information processing, generating commitment to outcomes to help people act morally, and facilitating abstract thought and communication. Other functions may be identified with more in-depth explorations from diverse disciplinary perspectives on the relationship between feelings and perceived risk.

Feelings that are integral to objects are often interpreted as signals of the value of those objects, motivating people to approach or avoid accordingly. Assessments of value based on integral affect differ from cognitive assessments in that the feelings tend to be more categorical, reference dependent, and sensitive to vivid imagery. Consequently, judgments based on integral feelings may be insensitive to scale (probability or magnitude) and myopic, emphasizing immediate hedonic consequences (positive or negative) over future consequences. The influence of specific characteristics of feelings (e.g., valence, intensity) on judgment processes needs further investigation.

Milder incidental feelings that are unrelated to the judgment target are also influential in judgment processes. In seeking information to inform their judgments, people tend to use whatever is available to them at the time and sometimes misattribute their mood states or metacognitive experiences as a reaction to the target. A variety of interventions can help people discern the diagnostic value of feelings.

## Further Research

---

Since empirical studies are designed in a specific theoretical and methodological context, no single study can fully answer the complex question of how feelings affect risk perceptions. However, to address the fragmented and sometimes inconsistent findings reported to date, future research needs to work to provide converging evidence for the role of feelings in judgment and decision processes. Converging evidence will be obtained by looking at multiple dependent variables and by using multiple methodological approaches to test alternative explanations of results (Weber and Hsee 1999). Though methods and measures for studying affect may be unfamiliar to many risk researchers, a wealth of tools exist in diverse disciplines studying the form and function of feelings. An interdisciplinary effort including physiological, neurological, psychological, sociological, and other approaches can be used to examine the interplay of affective and analytic processes in risk judgments, to yield the fullest understanding of risk reactions.

Future research also needs to explore new (e.g., qualitative) understandings of how affective and analytic processes (and their interactions) are best represented. A growing body of ethicists and social scientists have criticized purely quantitative approaches as ill-equipped to reflect public conceptualizations of the complex, multidimensional, and often nonmonetary qualities of risks being faced (Stern and Dietz 1994; Prior 1998; Satterfield and Slovic 2004; Finucane and Satterfield 2005; Roeser 2010). Likewise, the seemingly categorical, reference-dependent nature of the affective system may require new approaches to fully explicate nonintuitive consequences of feelings on risk judgments.

Another direction for future research is to evaluate the ecological validity of feelings. Adopting a Brunswikian (Brunswik 1952) approach, Pham (2004) suggests examining (a) the correlation between integral feelings elicited by objects and these objects' true criterion value (the ecological validity of feelings) and (b) the correlation between other available proxies of value and the object's criterion value (the ecological validity of alternative bases of evaluation). The ecological validity of incidental feelings could be examined in a similar fashion.

Finally, in a more practical realm, future research needs to help risk communicators and risk managers to determine the most effective tools for presenting and processing risk information. For instance, research will help to make risk estimates more accurate and risk mitigation behaviors more timely if it informs us of how to make abstract probabilities meaningful, reduce the gap between anticipated and experienced affect, facilitate the integration of non-commensurate metrics, or engage ethical assessments. Practitioners from diverse fields such as health care services, food safety, terrorism prevention, environmental resource management, and disaster preparedness would benefit from a systematic translation of the rich body of research into practice. Tools that account for the role of feelings in a way that facilitates efficient yet sound decision making will enhance our ability to successfully regulate risks.

## References

- Baron J (1997) Confusion of relative and absolute risk in valuation. *J Risk Uncertainty* 14(3):301–309
- Baron RS, Inman MB et al (1992) Emotion and superficial social processing. *Motiv Emotion* 16:323–345
- Bechara A, Damasio A (2005) The somatic marker hypothesis: a neural theory of economic decision. *Games Econ Behav* 52:336–372
- Bechara A, Damasio AR et al (1994) Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition* 50:7–15
- Bechara A, Damasio H et al (1997) Decision advantageously before knowing the advantageous strategy. *Science* 275:1293–1294
- Bell DE (1982) Regret in decision making under uncertainty. *Oper Res* 30:961–981
- Benthin A, Slovic P et al (1995) Adolescent health-threatening and health-enhancing behaviors: a study of word association and imagery. *J Adolesc Health* 17:143–152
- Berns GS, Chappelow J et al (2006) Neurobiological substrates of dread. *Science* 312:754
- Bodenhausen GV (1993) Emotions, arousal, and stereotype-based discrimination: a heuristic model of affect and stereotyping. *Affect, cognition, and stereotyping*. Academic, San Diego, pp 13–37
- Brunswik E (1952) The conceptual framework of psychology. University of Chicago Press, Chicago
- Cabanac M (1992) Pleasure: the common currency. *J Theor Biol* 155:173–200
- Clore GL, Huntsinger JR (2007) How emotions information judgment and regulate thought. *Trends Cogn Sci* 11:393–399
- Connolly T, Butler D (2006) Regret in economic and psychological theories of choice. *J Behav Decis Mak* 19:139–154
- Damasio AR (1994) *Descartes' error: emotion, reason, and the human brain*. Avon, New York
- Damasio A (1999) The feeling of what happens. Harcourt, Inc, New York
- Denes-Raj V, Epstein S (1994) Conflict between intuitive and rational processing: when people behave against their better judgment. *J Pers Soc Psychol* 66:819–829
- Epstein S (1994) Integration of the cognitive and the psychodynamic unconscious. *Am Psychol* 49:709–724
- Fellows LK, Farah MJ (2005) Different underlying impairments in decision making following ventromedial and dorsolateral frontal lobe damage in humans. *Cereb Cortex* 15:58–63
- Fetherstonhaugh D, Slovic P et al (1997) Insensitivity to the value of human life: a study of psychophysical numbing. *J Risk Uncertainty* 14(3):282–300
- Finucane M, Holup J (2006) Risk as value: combining affect and analysis in risk judgments. *J Risk Res* 9(2):141–164
- Finucane ML, Satterfield T (2005) Risk as narrative values: a theoretical framework for facilitating the biotechnology debate. *Int J Biotechnol* 7(1–3):128–146
- Finucane ML, Alhakami A et al (2000a) The affect heuristic in judgments of risks and benefits. *J Behav Decis Mak* 13:1–17
- Finucane ML, Slovic P et al (2000b) Public perception of the risk of blood transfusion. *Transfusion* 40:1017–1022
- Fischhoff B, Slovic P et al (1978) How safe is safe enough? A psychometric study of attitudes toward technological risks and benefits. *Policy Sci* 9:127–152
- Forgas JP (1992) Affect in social judgments and decisions: a multiprocess model. In: Zanna M (ed) *Advances in experimental social psychology*. Academic, San Diego, pp 227–275
- Fredrickson BL (1998) What good are positive emotions? *J Gen Psychol* 2:300–319
- Friedrich J, Barnes P et al (1999) Psychophysical numbing: when lives are valued less as the lives at risk increase. *J Consum Psychol* 8:277–299
- Frijda NH (1988) The laws of emotion. *Am Psychol* 43:349–358
- Frijda NH, Kuipers P et al (1989) Relations among emotion, appraisal, and emotional action readiness. *J Pers Soc Psychol* 57(2):212–228
- Gleicher F, Petty RE (1992) Expectations of reassurance influence the nature of fear-stimulated attitude change. *J Exp Soc Psychol* 28:86–100
- Hendrickx L, Vlek C et al (1989) Relative importance of scenario information and frequency information in the judgment of risk. *Acta Psychol* 72:41–63
- Higgins ET (1987) Self-discrepancy—a theory relating self and affect. *Psychol Rev* 94:319–340
- Hsee CK (1996) The evaluability hypothesis: an explanation for preference reversals between joint and separate evaluations of alternatives. *Organ Behav Hum Decis Process* 67(3):247–257
- Hsee C, Rottenstreich Y (2004) Music, pandas, and muggers: on the affective psychology of value. *J Exp Psychol* 133(1):23–30
- Isen AM (1997) Positive affect and decision making. In: Goldstein WM, Hogarth RM (eds) *Research on judgment and decision making: currents, connections, and controversies*. Cambridge University, New York, pp 509–534
- Isen AM (2000) Some perspectives on positive affect and self-regulation. *Psychol Inq* 11(3):184–187

- Isen AM, Geva N (1987) The influence of positive affect on acceptable level of risk: the person with a large canoe has a large worry. *Organ Behav Hum Decis Process* 39:145–154
- Jenkins-Smith H (2001) Modeling stigma: an empirical analysis of nuclear images of nevada. In: Flynn J, Slovic P, Kunreuther H (eds) *Risk, media and stigma*. Earthscan, London, pp 107–131
- Jenni KE, Loewenstein G (1997) Explaining the “identifiable victim effect”. *J Risk Uncertainty* 14(3):235–258
- Johnson EJ, Tversky A (1983) Affect, generalization, and the perception of risk. *J Pers Soc Psychol* 45:20–31
- Johnson M, Lakoff G (2002) Why cognitive linguistics requires embodied realism. *Cogn Linguistics* 13(3):245–263
- Kahan DM (2008) Two conceptions of emotion in risk regulation. *U Penn Law Rev* 156:741–766
- Kahan DM, Slovic P et al (2007) Affect, values, and nanotechnology risk perceptions: an experimental investigation. George Washington University Legal Studies, Washington, DC
- Kahneman D (2000) Experienced utility and objective happiness: a moment-based approach. In: Kahneman D, Tversky A (eds) *Choices, values, and frame*. Cambridge University Press and the Russell Sage Foundation, New York, pp 673–692
- Kahneman D (2003) A perspective on judgment and choice. *Am Psychol* 58(9):697–720
- Kogut T, Ritov I (2005) The “Identifiable Victim” effect: an identified group, or just a single individual? *J Behav Decis Mak* 18:157–167
- Lakoff G, Johnson M (1999) *Philosophy in the flesh*. Basic Books, New York
- Leiserowitz A (2006) Climate change risk perception and policy preferences: the role of affect, imagery, and values. *Clim Change* 77:45–72
- Lerner JS, Keltner D (2000) Beyond valence: toward a model of emotion-specific influences on judgment and choice. *Cogn Emotion* 14:473–493
- Lerner JS, Keltner D (2001) Fear, anger, and risk. *J Pers Soc Psychol* 81:146–159
- Lerner JS, Gonzalez RM et al (2003) Effects of fear and anger on perceived risks of terrorism: a national field experiment. *Psychol Sci* 14(2):144–150
- Loewenstein GF (1987) Anticipation and the valuation of delayed consumption. *Econ J* 97:666–684
- Loewenstein G, Lerner JS (2003) The role of affect in decision making. In: Davidson R, Goldsmith H, Scherer K (eds) *Handbook of affective science*. Oxford University Press, Oxford, pp 619–642
- Loewenstein GF, Weber EU et al (2001) Risk as feelings. *Psychol Bull* 127(2):267–286
- Loomes G, Sugden R (1982) Regret theory: an alternative theory of rational choice under uncertainty. *Econ J* 92:805–824
- Mackie DM, Worth LT (1989) Processing deficits and the mediation of positive affect in persuasion. *J Pers Soc Psychol* 57:27–40
- Maia TV, McClelland JL (2004) A reexamination of the evidence for the somatic marker hypothesis: what participants really know in the Iowa gambling ask. *Proc Natl Acad Sci* 101:16075–16080
- Meier BP, Robinson MD (2004) Why the sunny side is up. *Psychol Sci* 15(4):243–247
- Meier BP, Robinson MD et al (2004) Why good guys wear white: automatic inferences about stimulus valence based on color. *Psychol Sci* 15:82–87
- Mellers BA (2000) Choice and the relative pleasure of consequences. *Psychol Bull* 126(6):910–924
- Mowrer OH (1960) *Learning theory and behavior*. Wiley, New York
- Peters E (2006) The functions of affect in the construction of preferences. In: Lichtenstein S, Slovic P (eds) *The construction of preference*. Cambridge University Press, New York, pp 454–463
- Pfister H-R, Böhm G (2008) The multiplicity of emotions: a framework of emotional functions in decision making. *Judgem Decis Mak* 3(1):5–17
- Pham MT (2004) The logic of feeling. *J Consum Psychol* 14:360–369
- Pham MT (2007) Emotion and rationality: a critical review and interpretation of empirical evidence. *Rev Gen Psychol* 11(2):155–178
- Pham MT, Cohen JB et al (2001) Affect monitoring and the primacy of feelings in judgment. *J Consum Res* 28:167–188
- Prior M (1998) Economic valuation and environmental values. *Environ Values* 7:423–441
- Reyna VF (2004) How people make decisions that involve risk. *Curr Dir Psychol Sci* 13(2):60–66
- Reyna VF, Brainerd CJ (1995) Fuzzy-trace theory: an interim synthesis. *Learn Individ Differ* 7:1–75
- Reyna VF, Lloyd FJ et al (2003) Memory, development, and rationality: an integrative theory of judgment and decision-making. In: Schneider S, Shanteau J (eds) *Emerging perspectives on decision research*. Cambridge University Press, New York, pp 201–245
- Roeser S (2006) The role of emotions in judging the moral acceptability of risks. *Saf Sci* 44:689–700
- Roeser S (2010) Emotional reflection about risks. In: Roeser S (ed) *Emotions and risky technologies*. Springer, New York, 5, pp 231–244
- Roeser S et al (2009) The relation between cognition and affect in moral judgments about risk. In: Asveld L, Roeser S (eds) *The ethics of technological risks*. Earthscan, London, pp 181–201

- Rottenstreich Y, Hsee C (2001) Money, kisses, and electric shocks: on the affective psychology of risk. *Psychol Sci* 12(3):185–190
- Satterfield T, Slovic S (2004) What's nature worth? Narrative expressions of environmental values. University of Utah Press, Salt Lake City
- Satterfield T, Slovic P et al (2001) Risk lived, stigma experienced. In: Flynn J, Slovic P, Kunreuther H (eds) Risk, media and stigma. Earthscan, London, pp 68–83
- Schwarz N (1990) Feelings as information: informational and motivational functions of affective states. In: Sorrentino RM, Higgins ET (eds) Handbook of motivation and cognition: foundations of social behavior. Guilford, New York, pp 527–561
- Schwarz N (2004) Metacognitive experiences in consumer judgment and decision making. *J Consum Psychol* 14:332–348
- Schwarz N, Clore GL (1983) Mood, misattribution, and judgments of well-being: information and directive functions of affective states. *J Pers Soc Psychol* 45:513–523
- Schwarz N, Clore GL (1988) How do I feel about it? Informative functions of affective states. In: Fiedler K, Forgas J (eds) Affect, cognition, and social behavior. Hogrefe International, Toronto, pp 44–62
- Sloman SA (1996) The empirical case for two systems of reasoning. *Psychol Bull* 119(1):3–22
- Slovic P (ed) (2010) The feeling of risk: new perspectives on risk perception. Earthscan, London
- Slovic P, Flynn J et al (1991) Perceived risk, trust, and the politics of nuclear waste. *Science* 254:1603–1607
- Slovic P, Finucane ML et al (2002) The affect heuristic. In: Gilovich T, Griffin D, Kahneman D (eds) Intuitive judgment: heuristics and biases. Cambridge University Press, New York, pp 397–420
- Small DA, Loewenstein G (2005) The devil you know: the effects of identifiability on punitiveness. *J Behav Decis Mak* 18:311–318
- Small DA, Loewenstein G et al (2007) Sympathy and callousness: the impact of deliberative thought on donations to identifiable and statistical victims. *Organ Behav Hum Decis Process* 102:143–153
- Stern P, Dietz T (1994) The value basis of environmental concern. *J Soc Issues* 50(3):65–84
- Stevens SS (1975) Psychophysics. Wiley, New York
- Sunstein CR (2007) On the divergent American reactions to terrorism and climate change. *Colum L Rev* 50:503–557
- Trepel C, Fox CR et al (2005) Prospect theory on the brain? Toward a cognitive neuroscience of decision under risk. *Cogn Brain Res* 23:34–50
- Väistfjäll D, Peters E et al (2008) Affect, risk perception and future optimism after the tsunami disaster. *Judgem Decis Mak* 3:1
- von Neumann J, Morgenstern O (1947) Theory of games and economic behavior. Princeton University Press, Princeton
- Weber EU, Hsee CK (1999) Models and mosaics: investigating cross-cultural differences in risk perception and risk preference. *Psychon Bull Rev* 6(4):611–617
- Weber EU, Johnson EJ (2006) Constructing preferences from memory. In: Slovic P, Lichtenstein S (eds) Construction of preferences. Cambridge University Press, New York
- Wegener DT, Petty RE (1994) Mood management across affective states: the hedonic contingency hypothesis. *J Pers Soc Psychol* 66:1034–1048
- Wilson RS, Arvai JL (2006) When less is more: how affect influences preferences when comparing low- and high-risk options. *J Risk Res* 9(2):165–178
- Yeung CWM, Wyer RS (2004) Affect, appraisal, and consumer judgment. *J Consum Res* 31:412–424
- Zajonc RB (1980) Feeling and thinking: preferences need no inferences. *Am Psychol* 35:151–175
- Zeelenberg M, van Dijk WW et al (2000) On bad decisions and disconfirmed expectancies: the psychology of regret and disappointment. *Cogn Emotion* 14:521–541
- Zeelenberg M, Nelissen RMA et al (2008) On emotion specificity in decision making: why feeling is for doing. *Judgem Decis Mak* 3(1):18–27



# 27 Emotion, Warnings, and the Ethics of Risk Communication

Ross Buck<sup>1</sup> · Rebecca Ferrer<sup>2</sup>

<sup>1</sup>University of Connecticut, Storrs, CT, USA

<sup>2</sup>National Cancer Institute, Rockville, MD, USA

<b>Introduction</b> .....	<b>694</b>
<b>History</b> .....	<b>696</b>
Why Emotional Appeals Can Be Effective .....	696
Experienced Emotions and Decision-Making .....	697
Interactions of Emotion and Reason in Decision-Making .....	698
<b>Current Research</b> .....	<b>699</b>
Affective Versus Rational Modes of Cognition .....	699
Levels of Cognitive Processing in Judgment .....	699
Brain Lateralization and Style of Cognitive Processing in Judgment .....	702
Valence Versus Discrete Emotion Approaches .....	707
Valence Approaches .....	707
Discrete Emotion Approaches .....	708
Effective Public Health and Safety Appeals .....	711
Uncovering Specific Emotional Influences: An Application to Safer Sexual Behavior .....	712
<b>Further Research</b> .....	<b>715</b>
Emotion Intervention Strategies: Emotional Education and Emotional Competence .....	715
<b>Conclusion</b> .....	<b>717</b>

**Abstract:** In constructing warning messages, analytic-cognitive factors have traditionally been stressed: having script and/or images of sufficient size and legibility which show dangerous consequences and communicate how they can be avoided and safety maintained. Emotional, or affective-cognitive, factors have rarely been considered in the design of warnings. Indeed, employing emotional appeals to persuade an audience is widely regarded to be unethical, smacking of manipulation. However, emotional factors have been employed with great effectiveness in advertising and marketing, sometimes to actually weaken legitimate safety concerns. Advertisements for dangerous products routinely ignore risk or if required present warning information in a form that is easily overlooked or disregarded. However, emerging research on decision-making has found that emotion plays a critical role in reaching optimal conclusions. It is now recognized that emotional or affective-cognitive factors can influence judgment processes in many ways. First, there are immediate emotions involved in the decision itself, and the emotions anticipated to flow from the decision. Second, there are emotions intrinsic to the decision that may be evoked by the message itself, and emotions incidental to the decision that nevertheless may influence the outcome. Third, emotion can influence the degree of rationality or mindfulness in the deliberation itself. Effective warnings must command attention, stimulate memory, and evoke emotion, as well as communicate consequences and safe behavior. In addition, to construct effective warnings, one must recognize the emotions that people are likely to experience in dangerous situations, and help them to understand those feelings and desires in a context of mindfully managing risk.

## Introduction

---

Emotional appeals have historically been viewed as morally inferior to rational presentations of objective risk information. Emotional appeals are often regarded as manipulative and unethical, because they intentionally target processes that may be outside an individual's awareness in an attempt to produce a desired behavior (See Buck and Davis 2010). Emotional appeals have been strongly criticized as diminishing the “manipulee’s” abilities to make free and rational choices (Beauchamp 1988). Threat-based emotional appeals in particular have been described as explicitly using “the force of fear to try to manipulate human behavior” (Hastings et al. 2004), and have been denigrated for creating unnecessary consumer anxiety (Benet et al. 1993). These judgments persist even when critics acknowledge that social advertisers have good intentions (Arthur and Quester 2003). As such, officials often feel obligated to present risk information in a fashion that is as objective as possible, presenting only “factual” information about the numerical risks of a behavior, technology, or situation, and the potential benefits – again, framed in terms of change in objective risk – of a preventive behavior. This is intended to allow the public to make an autonomous, but informed, choice (see Buck and Davis 2010 for a detailed discussion). However, we argue that objectively presenting risk information outside of the context of emotion may not be the best, or morally preferred, strategy for two key reasons.

First, and importantly, emotional appeals have long been employed in the advertising field, where they can promote the mindless acceptance of risk (Buck and Davis 2010). For example, the popular advertising technique of *branding* involves using advertisements to encourage the public to associate emotionally charged images and videos with a certain consumer brand name. The intended outcome of this technique is essentially to increase positive emotions associated with the brand, which in turn directly and positively impacts purchase behavior

(e.g., Greifender et al. 2007; Pritchard and Morgan 1998). For example, through Nike's "Just Do It" campaign, consumers learn to associate with the Nike brand affectively laden images of fit, strong, and attractive elite athletes engaging in action-packed sporting activities, and the associated feelings of accomplishment, pride, and body satisfaction.

The consequences of these types of advertisements to the public may not be dramatic, beyond an increase in consumer spending on Nike products. However, emotional appeals in the media are not always so seemingly innocuous. In televised drug advertisements, the presentation of required information about significant side effects is often presented with images of health and happiness accompanied by soothing music, this emotional content directly contrasting with the information about risk presented often with rapid verbalization. Moreover, many advertising campaigns for tobacco and alcohol products promote unhealthy smoking or drinking behavior by linking their brands to affectively laden images associated with being fun-loving, popular, young, and loved. These types of advertisements have been utilized, with much success, by the tobacco and alcohol industry for decades, and they belie the small "warning labels" imposed on tobacco and alcohol products. In another example, fast food campaigns have successfully linked their brands with emotions associated with family and fun, such as loyalty, love, pleasure, and satisfaction. These emotional appeals in advertising are well-documented and quite effective in creating sustained changes in consumer behavior (e.g., Chaudhuri 2006), including increases in consumption of harmful products like cigarettes or unhealthy foods. These emotional appeals divert attention from potentially harmful effects of these behaviors, resulting in the mindless acceptance of risk on the part of consumers (Buck and Davis 2010).

Effective emotional risk messages by the public health field have the potential to counter these harmful emotional appeals from the advertising industry, and there are relevant examples. For example, cigarette packages in Canada and Brazil have been required to include large color photographs illustrating the harmful effects of smoking: preterm birth, impotence, tooth decay, cancer (Buck and Davis 2010). New US Food and Drug Administration regulations will similarly require such images to be introduced in the USA. However, risk communicators often miss the opportunity to make a difference in these behaviors by attempting to counter the advertising industry's emotional appeals with often ineffective appeals to reason. If emotional appeals are more effective than appeals to reason, it begs the question: Is it really more ethical to present objective and rational risk information when emotional appeals can promote behavior that is in the public's best interest? In a related line of reasoning, Thaler and Sustein (2008) argue that the choice environment and public policy affect decision-making at the individual level, and that policy makers have an obligation to structure choices in such a way that it "nudges" individuals toward decisions that are in their own best interest without forcing these choices.

A second reason that objectively presenting risk information may not be the morally preferred strategy is that, although emotional appeals in risk communication have been seen as unethical, it can be argued that all public health and safety appeals are emotional to some degree, since all perceptions and messages contain some affective component (Zajonc 1980). Indeed, Hilton (2008) argues that though the pragmatic view of communication calls for information to be presented in an objective way so that the decision maker may accurately interpret alternatives and make autonomous decisions, our language is structured in such a way that it always has evaluative properties, and even minor variances in sentence structure convey evaluation and emotion. This assertion is supported by examination of the effectiveness of warnings. A meta-analytic review of warnings indicated that, though warnings have a modest effect overall, there is considerable variability, with some

warnings producing behavior that is actually more unsafe than the behavior of the comparison group who does not receive a warning (Cox et al. 1997). Arguably, if individuals were behaving “rationally” and emotions were not aroused by warning messages, these warnings would either have a positive effect on behavior, or would have no effect if other considerations (e.g., pleasure in the risky behavior) trumped the risk information. However, the presence of these unintended boomerang effects of warnings indicates that there are some “irrational,” and perhaps emotional, effects of messages that are intended to be neutral. Thus, we argue that although emotion has been largely overlooked in the design of safety and public health risk messages, emotion is still present in these messages. If risk communicators always convey some emotion in their messages and warnings (Hilton 2008) and create an environment that predisposes certain choices (Thaler and Sustein 2008), is it truly unethical for them not to carefully target the emotional reactions they want to elicit based on empirical evidence in the field of emotion and decision-making?

Because it may be impossible to present information in a disinterested and unemotional manner, and because emotional appeals have the great potential to counter emotional appeals in the advertising industry that mindlessly promote the acceptance of risk, we submit that it is therefore the responsibility of health and safety officials to use emotions effectively to communicate information necessary for individuals to act in their own best interests. Here, we present support for the use of emotional appeals in warnings and risk messages, supporting the assertion that emotional appeals are not only ethical, but also essential in creating effective risk messages. The chapter reviews research concerning emotion and decision-making, and addresses neurobiological determinants of emotion and the implications for judgment and decision-making involving risk. We note the increasing appreciation of the role of emotion in decision-making and persuasion in recent years, after the long period of relative neglect in which rational information processing was emphasized. We conceptualize emotion in terms of affective cognition, and reason in terms of rational cognition, as detailed later in this chapter. The resulting “new look” in decision and persuasion theory has far-reaching implications for both theory and practice in the social and behavioral sciences. It also has the potential to bring academic theory and research more into line with approaches in the private sector, which has long used emotional appeals very effectively to steer decision-making, and indeed as noted at times deliberately to promote mindlessness and unwarranted risk-taking. We close the chapter with recommendations for future directions in emotional risk communication research, as well as recommendations for effective risk communication strategies that appeal to emotions in systematic and empirically driven ways.

## History

---

### Why Emotional Appeals Can Be Effective

---

It is now recognized that emotion, or affective cognition, can influence judgment processes in many ways. First, there are immediate emotions involved in the decision itself. For example, decisions involving risky sexual behavior can be influenced by arousal, excitement, and enthusiasm that may impede feelings of apprehension that might be anticipated were considerations of security and safety more salient. Second, there are emotions intrinsic to the

decision, and emotions incidental to the decision that nevertheless may influence the outcome. Unwise decisions about having risky sex may be greatly facilitated both by an attractive and engaging partner, and by a party atmosphere evoking the very excitement and arousal that motivate the risky behavior. Third, emotion may influence the degree of rationality or mindfulness in the deliberation itself. Here, we provide evidence for the crucial role of emotions in judgment and decision-making.

## Experienced Emotions and Decision-Making

As Damasio and colleagues have argued, emotion plays a critical role in adaptive judgment and decision-making (e.g., Damasio 1994; Bechara et al. 1997). One category of emotion that influences decision-making is “experienced emotion,” which can be both affective cognitions that are relevant to the decision (anticipatory emotions) and affective cognitions that are not directly relevant to the decision but influence it nonetheless (incidental emotions; Han et al. 2007; Loewenstein and Lerner 2003; Loewenstein et al. 2001). Anticipatory emotions are actually experienced at the time of the decision in response to cognitive representations of potential outcomes, and have been shown to influence behavior (Bagozzi et al. 1998). For example, when thinking about an upcoming vacation, one may actually *feel* excitement or contentment, rather than simply cognitively anticipating that these emotions may occur during the vacation in the future. Similarly, when thinking about being diagnosed with cancer, one may actually *feel* worry, dread, or anxiety, rather than simply cognitively anticipating that these emotions may occur during a hypothetical future diagnosis.

Incidental emotions may also be experienced at the time of the decision, but be unrelated to the decision; these also have a demonstrated relationship with behavioral decisions (e.g., Grunberg and Straub 1992; Harle and Sanfey 2007; Lerner and Keltner 2001; Lerner et al. 2004). For example, one may feel sad after watching a depressing movie, and though that sadness has no relationship to a decision about whether to purchase a product or engage in a particular health or safety-related behavior, it may influence that decision nonetheless. Though experienced emotions, both anticipatory and incidental, can be fleeting, their influence on decision-making can endure long after the emotional experience has ceased, perhaps because of the need to behavioral consistency (Andrande and Ariely 2009).

Additionally, anticipated emotions play a role in decision-making. These refer to cognitive conceptualizations of anticipated future emotions. Though anticipated emotions are not experienced at the time of the decision, they can contribute to the information that influences experienced anticipatory emotions, and can also contribute to decision-making at the cognitive level (Loewenstein and Lerner 2003; Loewenstein et al. 2001). For example, thinking about failing an examination in the future may conjure a cognitive judgment that the failure would be accompanied by disappointment; this anticipated emotion may then trigger actual experienced feelings of disappointment, or anticipatory disappointment. The cognitions of anticipated disappointment may be informed by past experience with failing a test. Similarly, imagining that one has been diagnosed with advanced cancer may conjure a cognitive judgment of anticipated regret of not being screened for cancer sooner; this anticipated regret may or may not trigger actual experienced emotions. In this way, past emotions may be used as a heuristic decision-making tool that informs perceptions of anticipated emotions, and indirectly, anticipatory emotions.

## Interactions of Emotion and Reason in Decision-Making

The influence of emotion on judgment and decision-making is important because, as discussed in the next section, affective cognitive reactions are often more rapid than rational cognitive reactions (e.g., LeDoux 1996; Zajonc 1980). These affective cognitive reactions can differ from rational cognitive reactions in a variety of complex ways (e.g., Ness and Klaas 1994; Lipkus et al. 2005; Loewenstein et al. 2001). These divergences are often resolved when the emotional reactions exert a dominant influence on judgment (e.g., Berridge and Aldridge 2008; Lawton et al. 2009; Norton et al. 2005). Additionally, rational cognitive and affective cognitive reactions can interact and influence each other in a variety of ways (Finucane et al. 2003), such as when anticipated emotions influence anticipatory emotions, as described above. Importantly, rational cognitive conceptualizations do not always influence affective reactions as might be expected. For example, affective reactions to risky consequences are largely unaffected by changes in the probability of those consequences, particularly when the consequences are affectively laden (e.g., Rottenstreich and Hsee 2001).

Recent research has demonstrated that emotion and cognition interact in important and interesting ways, influencing behavior in a manner that is not always anticipated. For example, preliminary longitudinal research has demonstrated that, among individuals who worry about cancer, risk perceptions actually have a *negative* relationship with cancer-preventive behaviors such as quitting smoking (e.g., Klein et al. 2009). Additionally, emotion can interact with framing of risk messages to affect behavioral outcomes. For example, enthusiasm and anger may temper the relationship between framing risk information in terms of gains and risk-aversion, whereas distress may increase this relationship (Druckman and McDermott 2008). That is, among individuals experiencing either enthusiasm or anger, gain-framed messages elicited less risk aversion, whereas among individuals experiencing distress, gain-framed messages elicited more risk aversion. These effects held for both naturally experienced and experimentally induced emotion.

Experienced emotions, both anticipatory and incidental, have important influences on the processing of risk information and on decisions made under risky circumstances. For example, fear tends to promote risk aversion, while anger and happiness tend to promote risk-seeking behavior (e.g., Fischhoff et al. 2005; Johnson and Tversky 1983; Lerner and Keltner 2001; Raghunathan and Pham 1999). It has been argued that individual differences in risk-taking preferences are influenced by tendencies of those individuals' emotional reactions to risky options (Hsee and Weber 1997), which is interesting in light of the fact that there are reliable individual differences in reliance on experiential thinking and affective reactivity (Gasper and Clore 1998; Peters and Slovic 2000). Recent research indicates that risk information may be processed through more affective-cognitive, as opposed to analytic-cognitive, pathways (e.g., Loewenstein et al. 2001; Slovic et al. 2004; Weinstein 1989; Weinstein et al. 2007). Further, affective variables such as feeling of dread may influence public acceptance of risk (e.g., Alhakami and Slovic 1994; Fischhoff et al. 1978; Slovic 1987), or may infuse risk information with meaning (Slovic et al. 2004).

Additionally, emotions affect the way that any information is processed and recalled. For example, positive moods facilitate more heuristic thinking, whereas certain negative moods, such as sadness, facilitate more systematic thinking (e.g., Bless 2000; Clore et al. 2002; Fiedler 2000). Individuals also attend more carefully to mood-congruent information,

and are more likely to recall such information when they are in a mood-congruent state (Bower 1991). This highlights the potential for affect to direct individuals' attention to different risk-related information (Diefenbach et al. 2008), orient individuals in the presence of information (Finucane et al. 2003), and act as a perceptual lens for interpreting situations (Lerner et al. 2003).

In light of the evidence presented above, it follows that risk messages should address affective reactions directly, rather than simply attempting to intervene on cognitive reactions.

## Current Research

---

### Affective Versus Rational Modes of Cognition

---

One of the most widely accepted aspects of the new look in emotion and decision-making is the notion that reason and emotion involve different kinds of cognitive processing, or different sorts of knowledge of events. These differences can be approached in two ways: first in terms of levels of cognitive processing from initial perception/appraisal to higher-order understanding, attribution, and perspective-taking; and second in terms of holistic-syncretic versus sequential-analytic styles of cognitive processing. In both cases, the cognitive process in question can be related to specifiable and observable or potentially observable brain mechanisms. Levels of cognitive processing can be related to levels in the hierarchy of neurochemical systems in the brain that respond to events, and styles of cognitive processing can be related to right- versus left-sided mechanisms in the brain.

### Levels of Cognitive Processing in Judgment

*Initial appraisal.* The process whereby the personal relevance of an event is apprehended for good or ill has been termed *appraisal*. Magda Arnold (1960a, b) saw appraisal to be instinctive and immediate: "even before we can identify something we may like or dislike it... There seems to be an appraisal of the sensation itself, before the object is identified and appraised" (1960b, p. 36). She suggested that appraisal is based upon a multilevel *estimative system*, the most fundamental of which she termed the *subcortical estimative system*. Arnold's conceptualization of appraisal accords well with recent evidence on emotional factors in decision-making. However, although her notion has been highly influential and adopted in a general way, most came to see appraisal as a relatively complex higher-order cognitive process.

*Analytic-cognitive appraisal.* Appraisal was defined by Richard S. Lazarus as *the evaluation of the harmful or beneficial significance of some event*. He and his colleagues demonstrated that the apparent harmfulness or stressfulness of the same stimulus event may be manipulated by manipulating appraisal. Spiesman et al. (1964) showed that a gruesome film depicting an apparently painful circumcision-like ritual in an aboriginal tribe caused strong physiological arousal. However, arousal was lessened by sound tracks which promoted *intellectualization* (describing the scene dispassionately and technically from an anthropological point of view) or *denial* (telling subjects that the ritual, while painful at points, was part of a happy celebration on the occasion of coming into manhood). Moreover, Lazarus and Alfert (1964) demonstrated

that the stress could be cognitively “short-circuited” by giving the denial information beforehand: This condition produced significantly less physiological arousal in viewers than the denial commentary accompanying the film.

In this model, events are first evaluated by a cognitive process termed *primary appraisal*. Information about the circumcision ritual being actually a joyful occasion presumably altered the primary appraisal of subjects to the film. If positive or negative emotions are aroused, coping strategies are selected to deal with the event in a cognitive process termed *secondary appraisal*. Some coping strategies are *emotion-focused*, attempting to reduce distressing emotion or increase positive emotion sometimes by defense mechanisms such as denial and intellectualization. Other coping strategies are *problem-focused*, doing something actively to change the situation for the better.

A major point of contention relating to Lazarus’ analysis is whether appraisal can be instinctive and intuitive, as Arnold (1960) argued, or requires more complex cognitive processing. Lazarus was explicit that cognition is both necessary and sufficient for the occurrence of emotion: “*sufficient* means that thoughts are capable of producing emotions; *necessary* means that emotions cannot occur without some kind of thought” (1991b, p. 353. Italics in the original). However, a classic debate with Robert Zajonc provided evidence that emotions can signal goal relevance, rather than the reverse.

*The Zajonc–Lazarus debate.* Studies by Robert Zajonc and colleagues suggested that an individual could respond preferentially to stimuli without knowing what they were. Using ambiguous stimuli like nonsense syllables and oriental ideograms, they showed that mere exposure to these stimuli produced liking. Persons expressed more liking the more that they were exposed to the stimuli, even though they could not consciously recognize them as being more familiar (Kunst-Wilson and Zajonc 1980; Wilson 1979). On this and other evidence, Zajonc suggested that affect occurs prior to, and independently of, cognition (Zajonc 1980, 1984). This ignited a classic debate with Lazarus (1982a, b).

Examining the arguments advanced by Zajonc and Lazarus reveals that their disagreement rests upon what each defines as “cognition.” Lazarus acknowledged that the comprehension that one’s well-being was at stake could be a “primitive evaluative perception” that may be “global and spherical.” In contrast, Zajonc required that cognition involve some kind of transformation of sensory input – some kind of “mental work” – and he complained that “Lazarus has broadened the definition of cognitive appraisal to include even the most primitive forms of sensory excitation” (1984, p. 117). Thus, Zajonc and Lazarus agreed that some sort of sensory information is required for emotion, but they disagree about what would constitute “cognition.”

*The LeDoux findings.* Joseph LeDoux and his colleagues demonstrated that there is direct sensory input to the amygdala and that this input is necessary for the learning of conditioned fear responses (LeDoux 1994). The amygdalae play a crucial role in responding to emotional events and storing emotional memories, as demonstrated by the Kluver-Bucy syndrome, which appeared when the amygdalae were removed bilaterally in monkeys (Kluver and Bucy 1939). Affected animals lost emotional responding even to innately feared stimuli such as snakes and fire; the animal mouthed even distasteful and painful stimuli such as dirt, feces, rocks, and burning matches.

The classical auditory and visual pathways proceed from the cochlea to the auditory neocortex, and retina to the visual neocortex, respectively. LeDoux and colleagues discovered pathways in the auditory and visual systems that diverge and proceed directly to the amygdalae (LeDoux et al. 1984). Therefore, the amygdalae have their own subcortical sensory inputs.

These inputs appear to constitute evolutionarily primitive early warning systems that trigger fast responses to threatening stimuli essential in the quick “assignment of affective significance to sensory events” (LeDoux 1993, p. 110). LeDoux termed this fast but relatively undifferentiated response to events the “low road” to cognition. The amygdala inputs are several synapses shorter than those of the classical sensory systems and therefore offer “a temporal processing advantage at the expense of perceptual completeness” (p. 112). Microseconds after the initial sensory input is received, more completely processed input enters the amygdalae from the neocortex and other centers associated with higher-order cognition, so that we are able to consciously “know” what caused the emotional reaction: This slower but more differentiated process LeDoux termed the “high road” to cognition.

LeDoux’s research speaks to the fundamental issue of the nature of the brain’s initial response to events. It appears that low road emotional responding precedes and can guide high-road cognition. Moreover, stimulation, lesions, and chemical manipulations (e.g., by drugs) of the central nervous system can lead to apparently complete emotional experiences and goal-directed behavior with no high-level cognitive “reason” for the state. This is the case despite the fact that the responder may recognize that the feelings are inappropriate and be striving unsuccessfully to control them.

The LeDoux studies had an important influence on the Zajonc–Lazarus debate, for they demonstrated that the subcortical amygdalae receive basic information about sensory events before the neocortex. In response, Lazarus (1991a) accepted that there are in fact two levels of cognitive processing: one direct and immediate, the other involving information processing. Lazarus acknowledged what he termed an “automatic mode of meaning generation” (1991a, p. 155). He suggested that there are two sorts of appraisal: one rapid, “automatic, involuntary, and unconscious”; the other “time-consuming, deliberate, volitional, and conscious” (Lazarus 1991a, p. 188).

*Appraisal and the frontal cortex.* The notion that there are multiple levels of cognition can be related directly to different processing and memory systems in the brain, and indeed it accords well with Magda Arnold’s (1960a, b) notion of multilevel “estimative systems.” LeDoux (1994) and Panksepp (1994) distinguished levels of cognition based upon brain mechanisms: “cortico-cognitive” processes are based on the neocortex and hippocampus, and “emotional” processing involves the amygdalae.

A neocortical region involved in the processing of the incentive value of events that receives strong inputs from the amygdalae is the orbitofrontal cortex (OFC), in the front of the brain immediately above the eyes. It has long been known that damage to the prefrontal cortex (PFC) in general and the OFC in particular produces serious deficits in decision-making, emotional processing, and social skills, which may be attributable to an overall insensitivity to future consequences. Damasio and colleagues have suggested that the key to the impairment in patients with OFC damage is their inability to generate normal somatic responses to emotionally charged events. Damasio’s (1994) Somatic Marker Hypothesis states that the positive or negative incentive values associated with appraisal and decision-making are stored as somatic markers in the ventromedial prefrontal cortex (VMPFC), which includes the OFC. Activation of these markers produces bodily feelings that in turn contribute to decision-making. On the other hand, Rolls (1999) suggested that requiring involvement of peripheral somatic processes is unnecessary, and that brain activity in the OFC and amygdalae, and brain structures connected with them, is related to felt emotion directly. Others suggest that the influence of emotion extends beyond valence, or positive–negative reactions, and that specific

emotions – happiness, security, fear, anger, guilt, sex, love, pride, pity, nurturance, resentment, and others – can have specific effects upon judgment and decision-making. We shall return to this issue later in the chapter.

## **Brain Lateralization and Style of Cognitive Processing in Judgment**

Multiple levels of knowledge and decision-making, from basic sensory awareness to higher-order rational cognitive processing, are clearly related to a hierarchy of neurochemical systems objectively visible in brain anatomy. Another concrete aspect of brain anatomy with implications for decision-making is cerebral lateralization. The left and right hemispheres of the brain are different in many respects – in embryological origin, in microstructure, and in gross anatomy – and these differences have functional implications. Most obviously, the left hemisphere (LH) is associated with language in the vast majority of human beings. The functions of the right hemisphere (RH) are less well understood but are the subject of intensive current research.

Lateralization of function appears in vertebrates including fish, reptiles, birds, and mammals (Denenberg 1981, 1984). Species with eyes placed laterally tend to scan for predators with the left eye, indicating RH involvement; while conspecific vocalizations tend to be processed in the LH (Des Roches et al. 2008). In most humans, language is processed in the LH, with Broca's area in the anterior LH associated with language expression and Wernicke's area in the posterior LH associated with language comprehension. Indeed, the human brain is more lateralized than most other vertebrate brains because the size of the corpus callosum, which connects the left and right neocortices, increases greatly during evolution, while the size of the anterior commissure, which connects left and right paleocortical and subcortical regions, does not. For this reason, many paleocortical and subcortical regions of the human brain are more directly connected with the ipsilateral neocortex than they are with corresponding paleocortical and subcortical regions on the other side (Ross 1992).

*Holistic-syncretic versus sequential-analytic styles of cognitive processing.* Tucker (1981) summarized evidence that the LH and RH are associated with different styles of conceptualization. The RH is characterized by an ability to holistically integrate and synthesize analog information from a variety of sources into a form of nonverbal conceptualization termed *syncretic cognition*. In syncretic cognition, sensory, affective, and cognitive elements are fused into a global construct. Syncretic ideation is particularly suited to picking up the affective meaning of complexes of nonverbal information provided by facial expression and gesture, vocal prosody and tone, and body movement and posture. In contrast, the LH is associated with *analytic cognition*, which is characterized by linear and sequential cognitive operations that can logically and rationally differentiate and articulate concepts. Language provides a prime example of analytic cognition. The global, holistic, and nonverbal knowledge of the RH is compatible with emotional cognition, while the verbal, linear, and sequential knowledge of the LH is compatible with rational cognition.

*Brain lateralization and emotion.* There are two major theories of the brain lateralization of emotion. The right hemisphere hypothesis states that the RH is specialized for all emotion processing (Borod et al. 1998), while the valence hypothesis holds that the RH is specialized for negative and the LH for positive emotion (Ahern and Schwartz 1979; Bogen 1985; Davidson 1992). A variant of the valence hypothesis was proposed suggesting that the RH is specialized

for avoidance emotions and the LH for approach emotions (Davidson and Irwin 1999). This answered evidence of LH involvement in emotions such as anger, which although considered to be hedonically negative is considered an approach emotion (Harmon-Jones et al. 2010).

Examination of the literature reveals that the right hemisphere hypothesis appears to be correct for emotional *communication*, both expression and recognition aspects (Borod and Koff 1990; Borod et al. 1986; Etcoff 1989; Ross 1981, 1992; Silberman and Weingartner 1986; Strauss 1986). There is much evidence that systems in the anterior and posterior RH are involved in the display and recognition of emotion, respectively. Buck and Van Lear (2002) suggested that these play roles in *spontaneous emotional communication* that are analogous to roles Broca's and Wernicke's areas in the LH play in the expression and comprehension of intentional symbolic messages, notably language.

The valence hypothesis may be more relevant to the question of the brain loci of emotional *experience*. Despite considerable evidence favoring the valence hypothesis, there are data that are difficult to reconcile. For example, although the theory holds that the LH is involved in positively valenced emotion, there is evidence that disgust is represented in the left insula (Calder 2003; Calder et al. 2000; Straube et al. 2010). This may be compatible with the approach-avoidance version of the valence hypothesis as, like anger, disgust albeit hedonically negative has been described as an approach emotion, and indeed a prosocial emotion, as a common response to disgust is to invite others to experience the disgusting object for themselves (Rozin et al. 1993). On the other hand, evidence that orgasm is based in RH mechanisms (Janszky et al. 2002; Holstege et al. 2003) seems difficult for both the valence and approach-avoidance theory accounts. Ross, Homan, and Buck (1994) suggested an alternate to the valence hypothesis that may answer some of the problems of the original conceptualization. This was based upon a study of patients undergoing the Wada test, where a brain hemisphere is temporarily deactivated by sodium amobarbital in preparation for brain surgery. Before the operation, patients were asked to describe a life event they had experienced that gave rise to strong emotion. During the Wada test while the RH was deactivated, they were asked about the same event. After the operation, they again described the event. Although the RH inactivation did not change the factual content of the life event, eight of the ten patients showed evidence of minimizing or denying primary emotions such as fear or anger. In a few cases, the RH inactivation seemed to produce a change in emotion. A man who described himself as "angry and frustrated" at the inability of physicians to diagnose his condition described himself as "sorry for people that they had so much trouble finding out what was wrong" when the RH was inactivated. A woman who said that she was "mad and angry" at being teased for her epilepsy as a child, when the RH was inactivated stated that she was "embarrassed" at the abuse. When the RH functioning was restored, the patients denied being sorry or embarrassed and insisted that they had been angry.

Ross et al. (1994) suggested that the RH inactivation produced changes consistent with a change in *type* of emotion in addition to valence: specifically a change from selfish to prosocial emotion (Buck 2002). This was consistent with an observation by Buck and Duffy (1980) in LH and RH damaged patients in responding to emotionally loaded slides, which suggested that the RH might be associated with the spontaneous expression of emotion, while the LH is associated with learned display rules: expectations about how and when emotions should be expressed. Specifically, most patient groups accentuated positive displays and attenuated negative displays, as would be expected from display rules, while LH-damaged patients did not.

The Ross et al. (1994) results suggested an extension of this: that the RH is associated with basic emotions, both negative and positive (e.g., the right-lateralization of orgasm), while the LH is associated with social emotions, including the voluntary modulation of RH-mediated basic emotions via display rules (pseudospontaneous communication: Buck and Van Lear 2002). Indeed, social emotions may often appear to be positively valenced because display rules often (but not always) encourage the expression of cheerful and positive albeit perfidious displays that are at variance with the “true feelings” of the responder. For example, Davidson and Fox (1982) demonstrated that social smiling is associated with LH activation in human infants, and interpreted this as consistent with the association of the LH with positively valenced emotions. However, another explanation is that the infants’ social smiles reflect prosocial attachment emotions as opposed to “selfish” pleasure. Indeed, stimuli used to elicit positive emotion may often elicit prosocial emotions as well (e.g., pictures of cute children, a baby gorilla). Unfortunately, stimuli used in these studies sometimes are not sufficiently described to judge whether the positive slides are actually prosocial in nature (e.g., Balconi et al. 2009).

Another consideration relevant to the Ross et al. (1994) hypothesis is that the LH is clearly associated with language, so that linguistically learned and structured display rules may be associated with LH responding regardless of the nature of the emotion being controlled. Moreover, language itself is not necessarily unemotional. Human beings derive pleasure when language is used well and frustration when it is misused, and it makes sense that such “linguistic emotions” are associated both with sociality and with LH processing (Buck 1988, 1994). There are known to be significant connections between Wernicke’s area in the LH and underlying limbic system structures, which exist most clearly in human beings (LeDoux 1986) and this could be a mechanism by which emotional factors are involved directly in language.

While relatively few studies have directly evaluated the Buck and Duffy (1980) and Ross et al. (1994) hypothesis, Shamay-Tsoory et al. (2008) found support in two studies. In the first, they tested the ability of patients with lesions in left versus right PFC to recognize photographs of six basic emotions (happy, sad, afraid, angry, surprised, disgusted from Ekman and Friesen 1975) and seven complex social emotions (interested, worried, confident, fantasizing, preoccupied, friendly, suspicious from Baron-Cohen et al. 1997). Shamay-Tsoory et al. found that left PFC-damaged patients were significantly more impaired in recognizing complex social emotions, and right PFC-damaged patients were slightly albeit not significantly more impaired in recognizing basic emotions. In the second study, they showed pictures of eyes posing basic versus complex social emotions to normal persons in left or right visual field presentations. This found the RH to be significantly better at recognizing basic than complex social emotions, and the LH to be slightly more accurate in recognizing complex social emotions than basic emotions. Shamay-Tsoory et al. (2008) concluded, consistent with Ross et al. (1994), that there is a RH advantage in recognizing basic emotions and an LH advantage in recognizing complex social emotions. The findings of Prodan et al. (2001) are also consistent: They showed that upper facial displays which are more likely to reflect basic emotions are processed in the RH, while lower face displays which more likely reflect the moderating effects of display rules are processed in the LH.

Another study relevant to the lateralization of basic versus social emotions was conducted by Gur and colleagues (Gur et al. 1995), who studied resting metabolism in brain areas associated with emotion. Among other things, results indicated that metabolism was left-lateralized in the cingulate gyrus, and right-lateralized in most ventro-medial temporal

lobe regions of the limbic system and their projections in the basal ganglia. As the amygdalae in the temporal-limbic region are associated with self-preservation, while the cingulate is associated with species preservation in MacLean's (1993) analysis, the pattern of greater relative left-sided cingulate metabolism and greater relative right-sided temporal limbic system metabolism accords well with the Ross et al. (1994) suggestion that selfish emotions are right-lateralized and prosocial emotions left-lateralized.

In conclusion, the evidence regarding the nature of the differences between the LH and RH in emotion is suggestive but not definitive, but attention should be paid to the selfish versus prosocial nature of stimuli as well as their valence and approach/avoidance implications. Studies of the lateralization of the amygdalae and PFC may shed further light on this issue.

*Lateralization of the amygdalae.* There is evidence that the amygdalae are asymmetrical, both structurally (Szabo et al. 2001) and functionally (Baas et al. 2004; Phelps et al. 2001; Schneider et al. 1997). We saw that LeDoux and colleagues showed the amygdalae to be necessary in the conditioned fear response. Subsequent research suggested that the right and left amygdalae are differentially involved in fear. Coleman-Mesches and colleagues found that temporary inactivation of the right, but not the left, amygdala by drug microinjection disrupted the retention of passive avoidance responding in rats (Coleman-Mesches and McGaugh 1995a, b, c), and that inactivation of the right but not the left amygdala attenuated the response to a reduction in reward (Coleman-Mesches et al. 1996). Also, lesions to the right amygdala led to larger reductions in fear responses than lesions to the left amygdala (Baker and Kim 2004). In humans, right temporal lobectomies led to lessened ability to recall unpleasant emotional events, while left temporal lobectomies did not (Buchanan et al. 2006). These results suggest that the right amygdala may make a greater contribution than the left to the memory for aversive experience. Also, Smith et al. (2008) described a case of post-traumatic stress disorder (PTSD) in a patient with no left amygdala, and suggested that because PTSD survived in the absence of the left amygdala, the right amygdala normally plays a greater role in its symptoms: fear conditioning, modulating arousal and vigilance, and maintaining memory for emotional context.

The functions of the left amygdala are not well understood but are the subject of much recent interest on the part of researchers. In a meta-analysis, Baas et al. (2004) found that, in studies of emotional processing, the left amygdala is more often activated than the right, and it appears to be activated by positive as well as negative emotion. Hardee et al. (2008) demonstrated that the left more than the right amygdala discriminated between increases in eye white area that signaled fear versus a similar increase in white eye area associated with a lateral shift in gaze direction, suggesting that parallel mechanisms code for emotional face information. This result is consistent with suggestion by Markowitsch (1998) that the left amygdala encodes emotional information with a relatively greater affinity to detailed feature extraction and language, while the right amygdala responds to pictorial or image-related emotional information in a relatively fast, shallow, or gross response. This analysis is consistent with the syncretic cognition-RH versus analytic cognition-LH distinction considered previously (Tucker 1981).

There are suggestions that the left amygdala is involved in the functioning of a "social brain" whose functioning is compromised in Asperger Syndrome (AS), a core feature of which is impaired social and emotional cognition. Fine et al. (2001) reported a case study of a patient with early left amygdala damage who was diagnosed with AS in adulthood. This patient had severe impairment in the ability to represent mental states (theory of mind, or ToM) in line with the AS diagnoses, but showed no indication of executive function impairment, suggesting both a dissociation between ToM and executive functioning, and that the left amygdala may play

a role in the development of circuitry mediating ToM. Ashwin et al. (2007) found that an AS group showed relatively less activation to fearful faces in the left amygdala and left OFC compared to controls. In contrast, left amygdala hyperactivation in response to emotional faces has been found in Borderline Personality Disorder (BPD: Donegan et al. 2003; Koenigsberg et al. 2009). The contrast between the symptoms of AS versus BPD and left amygdala hypoactivation and hyperactivation may therefore be relevant to understanding the unique functions of the left amygdala.

Both AS and BPD are associated with disrupted interpersonal relations, but for markedly different reasons. AS differs from autistic disorder in that there are no clinically significant delays in language, cognitive development, curiosity, or adaptive behavior other than in social interaction. Major AS symptoms include impairment of social interaction and communication, including deficits in the use of nonverbal behaviors to regulate social interaction (e.g., eye-to-eye gaze, facial expression). There is a typical lack of empathy (awareness of others or their needs), a lack of social or emotional reciprocity, and a preference for solitary activities (American Psychiatric Association 1994). In short, AS (and left amygdala hypoactivation) is associated with something of an obliviousness to other persons.

The pattern of social behavior in BPD is in many respects opposite. Major symptoms of BPD include impulsivity and emotional instability which can involve episodes of intense dysphoria, irritability, anxiety, and inappropriate anger that is difficult to control (American Psychiatric Association 1994). The individual may make frantic efforts to avoid real or imagined separation, rejection, or abandonment; which can lead to unstable and intense interpersonal relationships that alternate between extremes of idealization and devaluation. This pattern of dysregulation may result from a kind of social and emotional hypervigilance (Donegan et al. 2003), which speculatively could be related to left amygdala hyperactivation.

*Lateralization of the prefrontal cortex.* We can continue to consider these issues as we turn to neocortical systems closely connected with the amygdala: the right- and left-sided prefrontal neocortex (PFC) and their subregions. We noted the importance of the PFC in decision-making, particularly, the VMPFC and OFC subregions of the PFC. There is evidence that, like the amygdala, the functions of the PFC are lateralized: We have seen that the amygdala is closely connected with the OFC, and that persons with AS showed relatively less activation to fearful faces in both the left amygdala and left OFC (Ashwin et al. 2007).

Shamay-Tsoory et al. (2008) noted that the Buck and Duffy (1980) analysis implies that, to develop display rules, the child has attained explicit theory of mind (ToM) skills to attribute mental states to others. Also, while emotion recognition has been associated with the right PFC, recognition tasks involving ToM skills are associated with the left PFC. From this, Shamay-Tsoory et al. suggested that the right PFC plays a role in mediating basic emotions and the left PFC has a unique role in complex social emotions.

*Sex differences in brain lateralization.* So, does the left amygdala function with the left PFC along a social-emotional obliviousness to hypervigilance dimension, with normal social behavior and emotional communication requiring a moderate level of functioning? Normal women and men are sometimes said to vary along something of a vigilance-to-obliviousness dimension when it comes to social and emotional communication, so it is of interest that there is evidence of sex differences in amygdala and PFC lateralization. Killgore et al. (2000) hypothesized a redistribution of cerebral functions from the amygdala to the PFC from childhood to adolescence, reflecting greater self-control over emotional behavior. They found that, with increasing age, females showed an increase in PFC relative to amygdala activation in the left hemisphere in

response to fearful faces, whereas males did not show a significant age-related change. (Killgore and Yurgelun-Todd 2001) found both sexes to have greater left amygdala activation in response to fearful faces, while happy faces produced greater right than left amygdala activation in males but not females. Moreover, in women, activity in the left, but not right amygdala has been found to predict subsequent memory for emotional stimuli while right amygdala activity predicts emotional memory in men (Cahill et al. 2004). Cahill (2005) suggested that, in processing emotional experiences, women generally use the left and men the right amygdala, which among other things helps women remember details and men the central ideas of events.

In conclusion, there appears to be considerable evidence that the selfish-prosocial hypothesis is a viable alternative to valence and approach/avoidance hypotheses of cerebral lateralization. Neither the valence hypothesis nor the approach-avoidance analysis seems compatible with the observed sex differences, or the evidence that AS and BPD relate respectively to left (but not right) amygdala hypo- and hyperactivation, or the right-lateralization of orgasm. The selfish-prosocial hypothesis is compatible with these observations and also allows them to be placed in a wider context of the neurochemical basis of socio-emotional functioning including basic sex differences that are explainable in terms of evolutionary theory, with women being more empathic and verbal in accord with their greater caregiving role.

## Valence Versus Discrete Emotion Approaches

---

### Valence Approaches

Valence approaches assume that the effects of emotion on decision-making involve a dimension of positivity-negativity, with positive emotions having similar influences on behavior generally opposed to the influences of negative emotions (Elster 1998). One application of the valence approach isForgas's Affect Infusion Model (AIM), which posits that positive or negative affective states or moods have a persistent influence on judgment and decision-making by influencing the kinds of information people attend to, interpret, and recall (Forgas 1995), as well as attitudes, values, and judgments (Forgas 1999) and attributions for behavior (Forgas et al. 1990). For example, research in this line indicates that those in good moods make optimistic judgments, and those in bad moods make pessimistic judgments (e.g., Schwarz and Clore 1983). Also, positive emotions encourage the use of heuristics or mental shortcuts, while negative emotions encourage more complex cognitive processing (Schwarz 2002).

As noted previously, the somatic marker hypothesis also states that decision-making is guided by emotions (Damasio 1994). This hypothesis is supported by the research that demonstrates that patients with neural abnormalities in areas that govern emotions and feelings demonstrate abnormalities in these emotion and feelings, which in turn results in severe impairment of judgment and decision-making (e.g., Bechara et al. 2000, 2001; Bechara 2004). For example, individuals who have damage in the VMPFC, which as we have seen is an area involved in anticipatory emotional responses to decisions, are unable to feel anticipatory negative emotions associated with a risky gambling decision; thus, these individuals are unable to adapt to a decision-making task that requires them to learn to choose from less risky decks during the gambling process (Bechara et al. 1997). Emotions are thus seen as essential tools that aid in "rational" decision-making, and individuals who are unable to feel normal emotions are thus incapable of making good decisions.

## Discrete Emotion Approaches

More recent research has focused on the role of specific emotions in judgment and decision-making. In contrast to valence theorists, these researchers posit that different emotions can have different influences on behavior, regardless of whether they are the same valence. One such theory, the Appraisal Tendency Framework, divides emotions on a series of appraisals, including valence, but also other appraisals such as approach-avoidance (Han et al. 2007; Lerner and Keltner 2000). Research in this line has demonstrated that anger and fear, both negatively valenced emotions, have opposite influences on risk processing, such that anger causes individuals to be risk-seeking, whereas fear causes individuals to be risk-averse (Lerner and Keltner 2001).

*Evolutionary theory:* Another discrete emotion approach to persuasion and decision-making involves evolutionary theory. This approach suggests that affective stimuli can arouse specific emotions which motivate thoughts and actions consistent with the evolutionary functions of the emotion in question (Keltner et al. 2006). The evolutionary approach predicts that the effectiveness of an emotional appeal will depend upon the fit between the particular emotion and the context: A particular emotional appeal may be effective in one context and ineffective or counter-effective in another.

As an example, Griskevicius and colleagues (2009) showed that two emotions with clear evolutionary functions – fear and romantic desire – have different effects when combined with two different persuasive tactics from Cialdini and Goldstein (2004). One tactic used the principle of social proof: that if many others are doing it, it must be good; the second used the principle of scarcity: if it is rare, it must be good. Social proof (#1 product in the country!) was effective following the fear appeal but not the romantic desire appeal, while in contrast the scarcity appeal (limited time offer!) was effective when combined with romantic desire, but backfired when combined with fear. With additional analyses, the results suggested that fear motivated participants to stick together, while romantic desire motivated independence. The authors concluded that discrete emotions serve qualitatively distinct functions that cannot be captured by affective valence alone, and that these functions must be considered in judging the effectiveness of emotional appeals (Griskevicius et al. 2010).

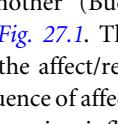
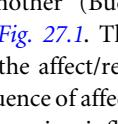
*Affective neuroscience.* A different approach, complementary to evolutionary theory, relates discrete emotions to specific neurochemical systems in the brain (Buck 1999). Paul D. MacLean (1993) was the originator of the Triune Theory of the brain and the limbic system concept, and a pioneer in the study of brain mechanisms of emotion. He suggested that there are three levels of processing systems in the brain: non-cortical (non-layered) *reptilian* systems characteristic of the brains of reptiles, old mammalian systems with 3–5 cortical layers (limbic system) characteristic of primitive mammals, and new mammalian systems with 6 layers (neocortex) characteristic of advanced animals including human beings. In addition, MacLean distinguished two general sorts of emotion at the level of the limbic system: some functioning in the survival of the individual and others functioning in the survival of the species. In his view, more complex species carry the older processing systems, which can influence and even unconsciously set the basic agenda for the newer systems.

Based upon MacLean's analysis, we suggest that there are in human beings *Reptilian* emotions involving “raw” sex and aggression (sex and power) based upon subcortical parts of the brain. Paleocortical (limbic system) areas are associated with more complex

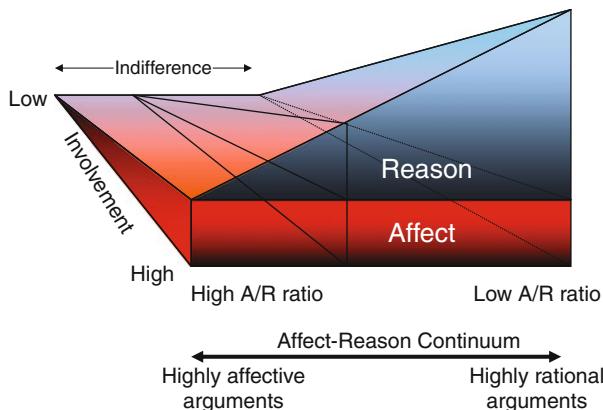
motivational-emotional systems. *Individualist* or selfish emotions involve self-preservation, and can be positive (happiness, satisfaction, security) and negative (anger, fear, sadness, disgust). *Prosocial* emotions involve species preservation, and can be positive (attachment, affiliation, love) and negative (isolation, loneliness, guilt, shame). There are also *Cognitive* emotions involved in the structuring of the cognitive system (curiosity, surprise, interest, boredom: Buck 1999). These specific emotions – reptilian, positive prosocial, negative prosocial, positive individualist, and negative individualist, are assessed by the Communication via Analytic and Syncretic Cognition (CASC) scale, which has been adapted to studies of the role of emotion versus reason in persuasion and social influence (Buck et al. 2004).

*The affect-reason-involvement (ARI) model.* A central assumption of this approach is that persuasion involves an interaction of affective and rational cognition, and this implies that involvement can be both rational and affective. Scales developed to measure involvement commonly include items appearing to assess both rational and affective involvement. For example, the McQuarrie and Munson (1987) revision of the Personal Involvement Inventory (RPI) includes items assessing both “risk” (no risk vs risky; easy to go wrong vs hard to go wrong; hard to pick vs easy to choose) and “hedonism” (appealing vs unappealing; unexciting vs exciting; fun vs not fun).

Chaudhuri and Buck (1995) defined *involvement* after Batra and Ray (1983) as the “depth and quality of cognitive processing” (p. 109), and invoked Tucker’s (1981) notion of syncretic versus analytic cognition to argue for two corresponding types of involvement. The tendency of a medium to encourage deep and high quality analytic processing defines its *rational involvement*, while its tendency to encourage deep and high quality syncretic processing defines its *emotional involvement* (Chaudhuri and Buck 1995; pp. 109–110. Italics in the original). Results of studies rating rational-analytic and affective-syncretic responses to 240 magazine and television advertising messages showed that, with a wide variety of product category and advertising strategy variables controlled, print media produced higher analytic-cognitive responses and electronic media higher syncretic-cognitive responses.

The differentiation of analytic and syncretic cognition blurs the common distinction between emotion and cognition: affect becomes a *type of cognition*, a type of knowledge, as we have seen. More specifically, affect involves syncretic cognition (feelings and desires), and reason involves analytic cognition, as defined previously (Tucker 1981). While affect and reason are often considered to be at ends of a continuum, we consider them to be qualitatively different kinds of systems that interact with one another (Buck 1985, 1988). The relationship between affect and reason is illustrated in  Fig. 27.1. The continuum at the base of  Fig. 27.1 describes the mix of affect and reason: the affect/reason continuum (A/R Continuum). On the extreme left of the continuum, the influence of affect is total: Reason has no influence. As one goes to the right, reason exerts an increasing influence relative to affect, although the influence of affect never falls to zero.

As noted, *involvement* is defined conceptually as the depth and quality of cognitive processing, and both affective and rational involvement are possible. Given this conceptualization, *Level of Involvement (LI)* can be defined operationally as the average of affective and rational involvement: that is:  $LI = (A + R)/2$ . In this way, involvement is defined both conceptually and operationally as a combination of affective and rational cognitive processing: If cognitive processing is measured, involvement is measured by definition. This suggests that one can be “hot, cold, or indifferent” in response to attitude objects and messages: “hot



**Fig. 27.1**

The affect-reason-involvement solid. Reason and affect interact on the face of the figure with reason having no influence on the left (high A/R ratio) and an increasing influence to the right (low A/R ratio). Involvement varies from a maximum on the face of the figure to low involvement (indifference) at the rear. An “ARI slice” is shown at the point where the influence of affect and reason are equal (Modified with kind permission from Fig. 2 in Buck et al. [2004])

processing” is relatively high in affect (high A/R Ratio) and high in involvement; “cold processing” is relatively high in reason (low A/R Ratio) and high in involvement; “indifference” is low in both affect and reason, and low in involvement.

The *ARI Solid* presented in [Fig. 27.1](#) models the relationships between affect, reason, and involvement. This is a three-dimensional figure bounded on one side by the A/R continuum and on the other by a low-high LI dimension. The relative influence of affect and reason at any point on the A/R continuum is represented by an *ARI Slice* in which the relative influence of affect and reason remains constant as involvement varies. The “floor” of the ARI solid is a two-dimensional space with the involvement dimension on the Y-axis and the A/R ratio on the X-axis (Buck et al. 2004). The position of an object (e.g., product or message) as represented by LI and the A/R ratio can be mapped on this surface, producing a Reason-Affect Map (RAM). For example, cars are rated high on both affect and reason: They are therefore in the middle of the A/R continuum and high in LI. Insurance is rated low on both affect and high on reason: It therefore is low on the A/R continuum and moderate in LI. Candy is rated high on affect and low on reason: It is therefore high on the A/R continuum and moderate in LI. Paper products are rated low on both affect and reason: They are therefore in the middle of the A/R continuum and low in LI. Messages – persuasive arguments – can also be rated for affect and reason and placed on the floor of the ARI solid.

*Operationalizing the interaction of discrete affects and reason.* Discrete emotion approaches imply that the measurement of affect can be highly specific, involving for example discrete reptilian, individualist, prosocial, and cognitive emotions. For example, Kowta and Buck (1995) asked college students from India and the USA to assess the rational involvement with buying a number of consumer products using standard questions from the McQuarrie and Munson (1987) Revised Personal Involvement Inventory. The students were also asked

what emotions were associated with buying the same products. For each of five kinds of emotion (curiosity, prosocial, individualist, reptilian sex, reptilian power), A/R ratios were computed for each product. This allowed the computation of Reason-Affect Profiles (RAPs) showing the relative mix of reason and affect across the specific emotions. Illustrative results are presented in Fig. 27.2. For condoms, all emotions save curiosity were rated higher than reason, suggesting relatively high affective and low rational involvement across the board. For greeting cards, only prosocial emotions were rated higher than reason. For both headache medicine and insect repellent, individualist emotions were rated as important, perhaps because the individual was seeking relief from pain or insects. Power was also rated high for the insect repellent, perhaps because the individual was using the product to kill insects. Also, interestingly, there were few differences between Indian and American college students in their ratings.

The results of the Kowta and Buck (1995) study are suggestive, indicating that persons are aware that emotional factors are involved in their decisions and able to report on their influence. The proof of the pudding, as it were, is to attempt such measurement, and to investigate whether a greater understanding of the specific emotions involved in behavior, including risky behavior, can be used to design more successful intervention programs aimed at changing such behaviors.

### Effective Public Health and Safety Appeals

Because emotion plays such a fundamental role in judgment and decision-making, risk communicators should pay attention to the affective content of their warnings and messages,

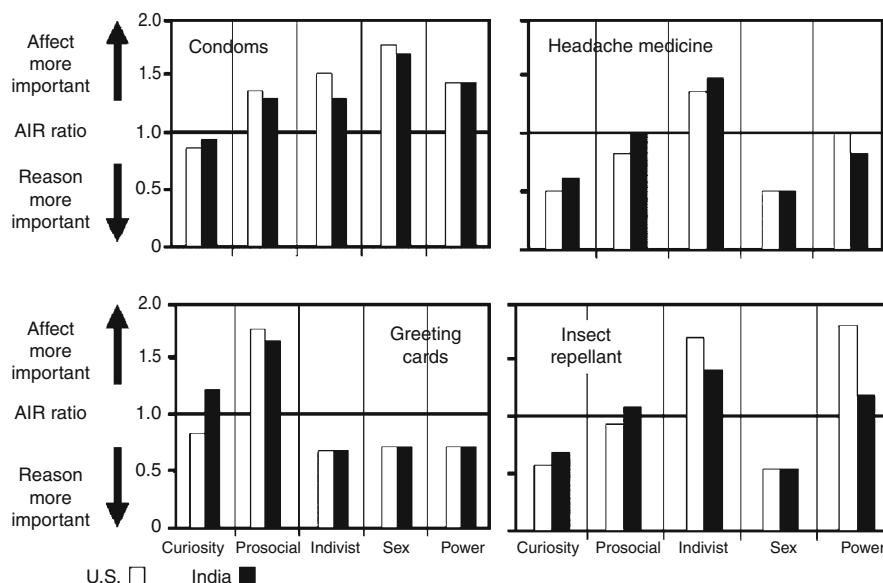


Fig. 27.2

Reason-affect profiles (RAPs) of four consumer products as rated by participants from America and India

and determine under what circumstances individuals are most likely to rely on their emotions to guide their decisions (Finucane 2008; Finucane et al. 2003), and what specific emotions are involved in such reliance. As stated, all communications, including public health and safety appeals, are infused with emotion, since all perceptions and messages contain an affective component (Hilton 2008; Zajonc 1980). Therefore, health and safety officials have an obligation and responsibility to the public to carefully attend to the emotional message they are presenting to the public and to design effective messages with emotion in mind. One of the essential aspects of such an enterprise is to understand the specific emotions involved in the behavior in question. A major example of risky behavior involving emotion is sex.

## **Uncovering Specific Emotional Influences: An Application to Safer Sexual Behavior**

Exploring the role of emotions in specific risky behaviors and their relation to and interaction with social-cognitive constructs may have meaningful implications for designing more effective risk messages. Prior to the design and implementation of emotion-based warnings and risk messages, researchers must first gain a more comprehensive understanding of the dynamics of emotions that are involved in specific risky situations and decisions. For example, if anxiety about a health risk could be reliably linked to adoption of preventive behavior, officials could confidently address anxiety in risk messages with an expectation that these types of appeals would be effective in facilitating appropriate behavior change.

*Emotions in sexual behavior: The SAFECOMM scale.* In an effort to understand the experience and dynamics of emotions involved in potentially risky sexual situations, Buck et al. (2004) developed a version of the CASC scale termed the Safe Sex Communication Scale (SAFECOMM). Participants were asked how “people feel” in a variety of situations including discussing condom use with a potential partner and having sex with and without condoms in situations differing in intimacy (one-night stand, friend, committed relationship). It was expected that condom nonuse would be related to a variety of negative emotions but also some positive emotions involving reptilian sex and power (the “reptilian rewards” hypothesis). Also, it was expected that there would be sex differences in ratings of anger and power, with women relative to men reporting more power but less anger when condoms are used.

Results showed support for the expected pattern of positive and negative prosocial, positive and negative individualistic, and reptilian sex and power emotions; although angry and fearful negative individualistic emotions were distinguished. Main effects indicated that, overall, condom nonuse compared to use was associated with generally higher ratings on negative individualistic and prosocial emotions, and lower ratings of security and confidence.

Negative prosocial emotions (feeling ashamed, embarrassed, and guilty) showed an interesting pattern. When condoms were used, these emotions were at low levels across relationships, but when condoms were not used these emotions were at high levels in one-night stands, decreasing as the relationship became more exclusive. Negative prosocial emotions showed significant interactions between gender and condom use. Women reported higher levels of shame, embarrassment, and guilt relative to men when condoms were not used, and women reported relatively higher levels of shame and embarrassment than did men for less exclusive relationships.

Fearful individualistic emotions (feeling afraid, nervous, and uncomfortable) showed a pattern similar to those of the negative prosocial emotions. Women reported higher overall feelings of nervousness and discomfort, but not of fear. Fear showed a significant interaction between gender and condom use: Women reported less fear than men did when condoms were used, but more when condoms were not used.

Angry individualistic emotions (feeling angry, unsure, insulted, and selfish) showed a pattern generally similar to that of the negative prosocial emotions, with highly significant interactions with relationship exclusivity. When condoms were used, these emotions were at relatively low levels, but when condoms were not used, they showed high levels in one-night stands that decreased as the relationship became more exclusive. Again, there were significant interactions with condom use: Women relative to men reported higher levels of anger, and being unsure and insulted, when condoms were not used. Also, women reported relatively higher levels of anger and being unsure in the less exclusive relationships.

Positive individualistic emotions (feeling secure, confident, and satisfied) generally showed a mirror image of the pattern shown with negative prosocial emotions. Significant main effects indicated that security and confidence were stronger when condoms were used, but the main effect for satisfaction was not significant, and indeed there was a tendency in the opposite direction, due to males' greater reported satisfaction associated with condom nonuse. Significant interactions indicated that, when condoms were used, these emotions were at relatively high levels, but when condoms were not used, they showed low levels in one-night stands that increased as the relationship became more exclusive. Also, positive individualistic emotions showed significant interactions with gender. Relative to men, women reported relatively lower levels of confidence and satisfaction when condoms were not used, and women reported relatively lower levels of security and satisfaction in less exclusive relationships.

Positive prosocial emotions (feeling loving/loved, caring, and intimate) increased as relationships became more exclusive. Interestingly, and importantly, in both sexes condom nonuse was associated with *lower ratings of caring but higher ratings of intimacy*. In the less exclusive relationships, these emotions were at roughly equivalent levels whether condoms were used or not used, but in the long-term relationship these emotions were consistently higher, for both sexes, when condoms were not used. Positive prosocial emotions showed no significant interactions between gender and condom use: Both for females and males, loving, caring, and intimacy tended to be higher in the long-term relationship when condoms were not used. For the less exclusive relationships, women reported relatively lower levels of loving/loved and caring relative to men.

Reptilian erotic emotions increased as relationships became more exclusive; erotic feelings were stronger when condoms were not used, and males reported generally higher erotic feelings. Ratings of erotic feelings showed no significant interactions between gender and condom use or gender and relationship exclusivity. Overall, reptilian power emotions increased slightly as relationships became more exclusive, and feelings of power were marginally stronger when condoms were not used. As expected, power ratings showed a significant interaction between gender and condom use/nonuse: Women indicated relatively greater power when condoms were used and men indicated relatively greater power when condoms were not used. Also as expected, men reported more anger than women when a condom was used and women reported being more angry when a condom was not used.

To summarize, positive emotions were high and negative emotions low when condoms are used; these emotions vary widely with the exclusivity of the relationship when condoms are not used. Also, as expected from the reptilian rewards hypothesis, condom nonuse was associated with higher ratings on reptilian sex and power. Also as expected, when condoms are not used feelings of power are higher for men than women and feelings of anger are higher for women than men. The strong relationships between reported emotions, relationship exclusivity, and condom use/nonuse are consistent with the notion that emotional variables exert important influences on decisions to use or not use condoms. Also, the results demonstrate the intricacy and subtlety of the influence of specific emotions, including reptilian and prosocial emotions not often recognized in many contemporary emotion theories. The complex results concerning positive prosocial emotions – that condom use as opposed to nonuse was seen to be associated positively with caring feelings but negatively with intimacy – was not expected, but on reflection is understandable.

The present authors replicated the Buck et al. (2004) study in a previously unpublished study which examined the hypothesized emotions involved in risky sexual situations, as well as cognitions involved in the same situations, allowing the ARI model to be employed in the analysis. Participants were 337 students from a large Northeastern US University who completed an online and expanded version of the SAFECOMM scale in response to a diverse set of sexual situations (described above). They were asked to rate how “most people would feel” in seven scenarios: discussing condom use with a new sexual partner; and using or not using a condom within sexual relationships of three levels of exclusivity (one-night stand, acquaintance, long-term relationship). The reliabilities of the emotion scales in this study were high ( $\alpha = 0.84\text{--}0.95$ ). Rational processing was operationalized by four questions about whether people think about using condoms (Do most people consider pros and cons of using a condom or not?) and three questions about whether people consider health consequences of using condoms (Do most people think about health consequences of using or not using a condom?) in each of the four situations (discussing, one-night stand, acquaintance, committed relationship). To obtain the A/R score for a given emotion, the emotion rating was divided by the mean of the reason items for each participant in that situation.

Highlights from our findings include the fact that positive prosocial and individualistic emotions increased as relationships become more committed, and negative prosocial and individualistic emotions decreased. Also, there were indications of relationship  $\times$  condom use interactions, such that increasing relationship exclusivity mitigated negative feelings associated with condom nonuse. But, this is not simply a matter of positivity and negativity: Reptilian sexual emotions were rated higher in committed relationships when condoms are NOT used, illustrating again the “reptilian rewards” of not using condoms. Regarding the A/R scores, or emotion *relative to reason*, discussing condom use was relatively the most “rational” of the situations, with reason predominating throughout. The most emotional situation was clearly the committed relationship, with positive prosocial, positive individualistic, reptilian power, and reptilian sex all rated as stronger than reason in that situation. Also, it is noteworthy that reptilian sex was at relatively high levels across the situations.

Together these results provide preliminary evidence for the potential effectiveness of a risk communication strategy that targets positive prosocial and individualistic emotions in the context of committed relationships, and negative prosocial and individualistic emotions in the context of more casual relationships. Additionally, a risk communication strategy that addresses sexual emotions related to condom nonuse may be promising, considering the high

levels of reptilian sex emotions associated with all situations and in particular with situations where condoms are NOT used. Additionally, emotion-based risk communication strategies may prevail over more social or cognitive strategies when addressing actual condom use as opposed to discussing condoms with a new partner, since the A/R scores demonstrate the most “rational” thinking in relation to discussions as opposed to actual sexually charged situations.

## Further Research

---

### Emotion Intervention Strategies: Emotional Education and Emotional Competence

---

One way to address emotions in risk communication is through *emotional education* (Buck 1983, 1985, 1990, 1994). Emotional education is a concept introduced in the Developmental Interactionist (DI) Theory of Emotion, which posits an interaction between reason and emotion that occurs during the course of normal individual development (Buck 1985, 1988, 1999, 2002). DI theory states that individuals learn how to label and understand feelings in childhood by interpreting others' responses to their emotional displays, resulting in naturally occurring emotional education. Basically, emotional education is the process of teaching individuals to correctly (or incorrectly) identify emotions they are experiencing, and what to do when they occur.

Though DI theory posits emotional education as a naturally occurring process, it has also been proposed as means of risk communication and behavior change intervention (e.g., Buck 1985, 1990; McWhirter 1995), and encouraged as a means of supplementing existing risk messages, such as sexual risk reduction interventions (Shaughnessy and Shakesby 1992). According to Buck (1990), though individuals have generalized access to their subjective emotional feelings, they may be unable to consciously identify the cause of the emotion or label the emotion itself. Emotional education in a risk communication strategy may help individuals to identify and label the emotions they experience. When an individual is able to identify experienced emotions confidently and reliably, and can deal with and express them effectively in a way that is appropriate to the situation and in the individual's best interest, this is considered *emotional competence*. Emotional competence is the desired outcome of any emotional education intervention (e.g., Buck 1985, 1990; Buck and Powers 2011; McWhirter 1995).

Research shows that individuals are relatively poor at anticipating how they will feel in a given situation, and how they will react to that emotion in the situation, when they are not emotional at the time of prediction (Finkenauer et al. 2007; Gilbert et al. 1998; Loewenstein 1996). Emotional education may be one way to help individuals to become better at anticipating their emotions, and in developing strategies in advance that might be helpful in dealing with these emotions in ways that facilitate more mindful and therefore safer sexual decision-making. For example, it may be possible to teach individuals to recognize and anticipate positive emotions associated with a healthy or safe behavior, or negative emotions associated with non-adherence to that behavior, and to either increase the experience and the anticipation of these emotions, as well as a strategic response to these emotions that results in the target behavior. If emotional education is successful, it could have a great impact on behavior: as stated above, anticipated and anticipatory emotions have both direct and indirect effects on judgment, decision-making, and behavior (e.g., Loewenstein and Lerner 2003).

In the advertising and marketing field, such an emotional-education approach is quite common. As described earlier, the popular advertising technique of branding involves educating individuals to associate particular affectively laden images with a certain brand, in order to increase emotions associated with the brand and strengthen the link between these emotions and purchase behavior (e.g., Greifender et al. 2007; Pritchard and Morgan 1998). Branding's effects on consumer behavior are well-documented (e.g., Buck and Davis 2010; Chaudhuri 2006).

Though most popular in advertising and marketing, emotional education strategies have been used preliminarily in the risk communication domain as well, though with far less prevalence than in the advertising industry. For example, recent research has directed individuals, through use of role-playing, to "pre-live" emotional consequences of positive and negative genetic testing results when making the decision whether or not to engage in such testing behavior (Diefenbach and Hamrick 2003; Miller et al. 2001). This pre-living of emotion can be seen as a form of emotional education.

One potentially efficacious method of facilitating emotional education and emotional competence is through the use of videos, as people often seek out media for emotional education in everyday life. Videos offer unique access to both the feelings of others as well as a viewer's own feelings by detailing situations in which actors are confronted with emotional stimuli and describe their physical and emotional responses (e.g., Buck 1988; Boyanowsky et al. 1972; Cantor 1982). Movies provide opportunities to consider situations that would provoke various emotional outcomes, as well as examples of others' responses to such situations. For example, Boyanowsky et al. (1972) found that, after a murder took place on campus, attendance at a movie depicting a murder increased, while attendance at a similar film with no murder remained steady, as compared to showings of the same two movies prior to the campus murder. Further, women who lived in the same dorm as the murdered woman were more likely to choose to attend the murder movie than the other movie. This could indicate that individuals may seek out videos in order to facilitate their own emotional education and competence after an emotional event occurs in real life. Thus, media and film are a natural way to facilitate emotional education and emotional competence in risk communication, and can be used in risk messages to affect systematic exposure to different affective-laden situations in a manner that would not otherwise be possible.

Indeed, recent research has demonstrated that videos designed to facilitate emotional education and emotional competence with relation to sexual situations significantly increased condom use when paired with traditional social-cognitive intervention material (Ferrer et al. 2011). The emotional education intervention content was designed to increase individuals' identification and anticipation of specific emotions and to facilitate enactment of response strategies that result in condom use. The design of the intervention drew on results of the SAFECOMM studies described above. The videos addressed sexual situations that occur both inside and outside the context of a committed relationship, and emotions identified by the SAFECOMM studies to be particularly relevant in each context. Positive prosocial and individualistic emotions, such as caring, intimacy, and confidence, were addressed in the context of committed relationships. Also, positive individualistic emotions were targeted in the context of more casual relationships, in addition to targeting negative prosocial and individualistic emotions such as embarrassment, guilt, and detachment. Sexual reptilian emotions were addressed in both relationship contexts, as these were associated with all situations and in particular in situations where condoms were *not* used. All participants

watched videos about both relationship contexts, as relationships among college students can be transient and as such, these individuals may quickly move from a committed to a casual relationship status (or vice versa).

Participants watched the videos designed to highlight these relevant emotions: love and caring for a partner that can be expressed with condom use; confidence and security in a sexual situation as a result of condom use; embarrassment or guilt associated with not using a condom; and eroticism that could be achieved even while using a condom. Discussions followed the videos to reinforce the emotional education. Individuals were randomly assigned, in small groups, to receive (1) the emotional education intervention paired with traditional social-cognitive intervention material, (2) the social-cognitive intervention material only, or (3) no intervention. Compared to the control condition, both the intervention conditions reported increased condom use at 3 months post-intervention. However, at 6 months post-intervention, only the group who had received the emotional education intervention sustained changes in condom use behavior. This study provides preliminary evidence suggesting that emotional appeals to health promotion behavior may be more efficacious in sustaining behavior change than more “rational” appeals alone. Additionally, the intervention was one of few single session interventions demonstrated to increase and sustain condom use, indicating the powerful potential of emotional appeals in effecting desired behavior change. Additional research is necessary to replicate these findings and to demonstrate their efficacy in other health and safety domains; however, these findings provide convincing evidence that such strategies have promise and should be examined further as risk communication strategies that could be used effectively by public health and safety officials.

One essential direction for future research concerns evaluating the effectiveness of emotional appeals in the “real world.” Although such appeals have been demonstrated to change behavior in laboratory and more controlled settings, there is no definitive empirical evidence to demonstrate what types of emotional appeals are most likely to be effected in “the sophisticated and overcrowded clutter of the real-world communications environment” (Hastings et al. 2004). We also know little empirically about the long-term exposure to repeated emotional appeals (Hastings et al. 2004). The marketing and advertising industry has certainly demonstrated that repeated emotional appeals distributed en masse to the public can be utilized effectively (e.g., Chaudhuri 2006). However, publicly available empirical evidence concerning the properties and characteristics of effective (and ineffective) appeals at the mass communication level is sparse. Additional research is necessary to determine what types of emotional appeals are most likely to translate from a controlled environment to the real world.

## Conclusion

In conclusion, the Ferrer et al. (2011) study illustrates the potential of incorporating explicit emotional education appeals in risk messages and interventions. By first understanding the specific emotions that are likely to be encountered in a risky situation, and then by presenting information on how to understand and deal with those emotions, it becomes possible to alert vulnerable persons to anticipate their feelings and desires in ways that promote mindful choice: to appreciate, for example, that sexually transmitted diseases do not “stay in Vegas.” The promotion of mindful choice is clearly beneficial and ethical, even though it may at times

involve manipulating emotion. The full potential of such warning tactics and strategies have yet to be fully tested and evaluated, but they are based upon both the latest research evidence and the compelling and sobering long-term example of success of advertising and marketing.

## References

- Ahern G, Schwartz G (1979) Differential lateralization for positive versus negative emotion. *Neuropsychologia* 17:693–698
- Alhakami AS, Slovic P (1994) A psychological study of the inverse relationship between perceived risk and perceived benefit. *Risk Anal* 14:1085–1096
- American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th edn, revised. American Psychiatric Association, Washington, DC
- Andrade EB, Ariely D (2009) The enduring impact of transient emotions on decision making. *Organ Behav Hum Decis Process* 109:1–8
- Arnold M (1960) Emotions and personality, vol 1 and 2. Columbia University Press, New York
- Arthur D, Quester P (2003) The ethicality of using fear for social advertising. *Australas Mark J* 11:12–27
- Ashwin C, Baron-Cohen S, Wheelwright S, O'Riordan M, Bullmore ET (2007) Differential activation of the amygdala and the 'social brain' during fearful face-processing in Asperger Syndrome. *Neuropsychologia* 45:2–14
- Baas D, Aleman A, Kahn RS (2004) Lateralization of amygdala activation: a systematic review of functional neuroimaging studies. *Brain Res Rev* 45:96–103
- Bagozzi RP, Baumgartner H, Pieters R (1998) Goal-directed emotions. *Cogn Emotion* 12:1–26
- Baker KB, Kim JJ (2004) Amygdala lateralization in fear conditioning: evidence for greater involvements of the right amygdala. *Behav Neurosci* 118:15–23
- Balconi M, Falbo L, Brambilla E (2009) BIS/BAS responses to emotional cues: self report, autonomic measure, and alpha brain modulation. *Pers Individ Differ* 47:858–863
- Baron-Cohen S, Jolliffe T, Mortimore C, Robertson M (1997) Another advanced test of theory of mind: evidence from very high functioning adults with autism or Asperger syndrome. *J Psychol Psychiatry Allied Discip* 38:812–822
- Batra R, Ray ML (1983) Conceptualizing involvement as depth and quality of cognitive response. In: Bagozzi RP, Tybout AM (eds) Advances in consumer research, vol 10. Association for Consumer Research, Ann Arbor, pp 309–313
- Beauchamp TL (1988) Manipulative advertising. In: Beauchamp TL, Bowie NE (eds) Ethical theory and business, 3rd edn. Prentice Hall, Englewood Cliffs
- Bechara A (2004) The role of emotion in decision-making: evidence from neurological patients with orbitofrontal damage. *Brain Cogn* 55:30–40
- Bechara A, Damasio H, Tranel D, Damasio AR (1997) Deciding advantageously before knowing the advantageous situation. *Science* 285:1293–1294
- Bechara A, Damasio H, Damasio AR (2000) Emotion, decision-making, and the orbitofrontal cortex. *Cereb Cortex* 10:295–307
- Bechara A, Damasio H, Tranel D, Anderson SW (2001) Dissociation of working memory from decision making within the human prefrontal cortex. *J Neurosci* 18:428–437
- Benet S, Pitts RE, LaTour M (1993) The appropriateness of fear appeal use for health-care marketing to the elderly – is it OK to scare granny? *J Bus Ethics* 12:45–55
- Berridge KC, Aldridge JW (2008) Decision utility, the brain, and the pursuit of hedonic goals. *Soc Cogn* 26:621–646
- Bless H (2000) Moods and general knowledge structures: happy moods and their impact on information processing. In:Forgas JP (ed) Feeling and thinking: the role of affect in social cognition. Cambridge University Press, New York, pp 131–142
- Bogen JE (1985) The callosal syndromes. In: Heilman KM, Valenstein E (eds) Brain mechanisms underlying speech and language. Grune and Stratton, New York
- Borod J, Koff E (1990) Lateralization of facial emotional behavior: a methodological perspective. *Int J Psychol* 25:157–177
- Borod J, Koff E, Buck R (1986) The neuropsychology of facial expression: data from normal and brain-damaged adults. In: Blanck P, Buck R, Rosenthal R (eds) Nonverbal communication in the clinical context. Penn State University Press, University Park
- Borod JC, Cicero BA, Obler LK, Welkowitz J, Erhan HM, Santschi C, Grunwald IS, Agosti RM, Whalen JR (1998) Right hemisphere emotional perception: evidence across multiple channels. *Neuropsychology* 12(3):446–458

- Bower GH (1991) Mood congruity of social judgments. In:Forgas JP (ed) Emotion and social judgments. Pergamon, New York, pp 31–53
- Boyanowsky EO, Newton D, Walster E (1972) Film preferences following a murder. *Commun Res* 1:32–43
- Buchanan TW, Tranel D, Adolphs R (2006) Memories for emotional autobiographical events following unilateral damage to medial temporal lobe. *Brain* 129:115–127
- Buck R (1983) Emotional development and emotional education. In: Plutchik R, Kellerman H (eds) Emotion in early development. Academic, New York
- Buck R (1985) Prime theory: an integrated view of motivation and emotion. *Psychol Rev* 92:389–413
- Buck R (1988) Human motivation and emotion, 2nd edn. Wiley, New York
- Buck R (1990) Rapport, emotional education, and emotional competence. *Psychol Inq* 1:301–302
- Buck R (1994) The neuropsychology of communication: spontaneous and symbolic aspects. *J Pragmatics* 22:265–278
- Buck R (1999) The biological affects: a typology. *Psychol Rev* 106:301–336
- Buck R (2002) The genetics and biology of true love: prosocial biological affects and the left hemisphere. *Psychol Rev* 109:739–744
- Buck R, Davis WA (2010) Marketing risk: emotional appeals can promote the mindless acceptance of risk. In: Roeser S (ed) Emotions and risky technologies. Springer, Berlin
- Buck R, Duffy RJ (1980) Nonverbal communication of affect in brain-damaged patients. *Cortex* 16:351–362
- Buck R, Powers SR (2011) Emotion, media, and the global village. In: Doveling K, von Scheve C, Konijn EA (eds) The routledge handbook of emotions and mass media. Routledge, Abingdon, Oxon, New York
- Buck R, Van Lear CA (2002) Verbal and nonverbal communication: distinguishing symbolic, spontaneous, and pseudo-spontaneous nonverbal behavior. *J Commun* 52:522–541
- Buck R, Anderson E, Chaudhuri A, Ray I (2004) Emotion and reason in persuasion: applying the ARI model and the CASC scale. *J Bus Res* 57:647–656
- Cahill L (2005) His brain, her brain. *Sci Am* 292(5):40–47
- Cahill L, Uncapher M, Kilpatrick L, Alkire MT, Turner J (2004) Sex-related hemispheric lateralization of amygdala function in emotionally influenced memory: an fMRI investigation. *learning and memory. Learn Mem* 11:261–266
- Calder AJ (2003) Disgust discussed. *Ann Neurol* 53:427–428
- Calder AJ, Keane J, Manes F, Antoun N, Yong AW (2000) Impaired recognition and experience of disgust following brain injury. *Nat Neurosci* 3:1077–1078
- Cantor J (1982) Developmental studies of children's fright from mass media. In: Paper presented at the convention of the International Communication Association, Dallas
- Chaudhuri A (2006) Emotion and reason in consumer behavior. Butterworth, Woburn
- Chaudhuri A, Buck R (1995) Media differences in rational and emotional responses to advertising. *J Broadcasting Electron Media* 39(1):109–125
- Cialdini RB, Goldstein NJ (2004) Social influence: compliance and conformity. *Annu Rev Psychol* 55:591–621
- Clore GL, Gasper K, Garvin E (2002) Affect as information. In: Forgas JP (ed) Handbook of affect and social cognition. Lawrence Erlbaum, Mahwah
- Coleman-Mesches K, McGaugh JL (1995a) Differential involvement of the right and left amygdala in expression of memory for aversively-motivated training. *Brain Res* 670:75–81
- Coleman-Mesches K, McGaugh JL (1995b) Muscimol injected into the right or left amygdaloid complex differentially affects retention performance following aversively motivated training. *Brain Res* 676:183–188
- Coleman-Mesches K, McGaugh JL (1995c) Muscimol injected into the right or left amygdaloid complex differentially affects retention performance following aversively motivated training. *Brain Res* 676:183–188
- Coleman-Mesches K, Salinas JA, McGaugh JL (1996) Unilateral amygdala inactivation after training attenuates memory for reduced reward. *Behav Brain Res* 77:175–108
- Cox EP, Wogalter MS, Stokes SL (1997) Do product warnings increase safe behavior? A meta-analysis. *J Public Policy Mark* 16:195–204
- Damasio A (1994) Decartes' error: emotion, reason, and the human brain. Gosset/Putnam, New York
- Davidson RJ (1992) Prolegomenon to the structure of emotions: Gleanings from neuropsychology. *Cogn Emot* 6:245–258
- Davidson RJ, Fox NA (1982) Asymmetrical brain activity discriminates between positive and negative affective stimuli in human infants. *Science* 218: 1235–1237
- Davidson RJ, Irwin W (1999) The functional neuroanatomy of emotion and affective style. *Trends Cogn Sci* 3:11–21
- Denenberg VH (1981) Hemispheric laterality in animals and the effects of early experience. *Behav Brain Sci* 4:1–19
- Denenberg VH (1984) Behavioral asymmetry. In: Geschwind N, Galaburda AM (eds) Cerebral dominance: the biological foundations. Harvard University Press, Cambridge, MA, pp 114–133

- Des Roches ADB, Richard-Yris M-A, Henry S, Ezzaouia M, Hausberger M (2008) Laterality and emotions: visual laterality in the domestic horse (*Equus caballus*) differs with objects' emotional value. *Physiol Behav* 94:487–490
- Diefenbach MA, Hamrick N (2003) Self-regulation and genetic testing. In: Cameron LD, Leventhal H (eds) *The self-regulation of health and illness behavior*. Routledge, London, pp 314–331
- Diefenbach MA, Miller SM, Porter M, Peters E, Stefanek M, Leventhal H (2008) Emotions and health behavior: a self-regulation perspective. In: Lewis M, Haviland-Jones JM, Feldman Barrett L (eds) *Handbook of emotions*. Guilford, New York
- Donegan NH et al (2003) Amygdala hyperreactivity in borderline personality disorder: implications for emotional dysregulation. *Biol Psychiatry* 54(11):1284–1293
- Druckman JN, McDermott R (2008) Emotion and the framing of risky choice. *Polit Behav* 30:297–321
- Ekman P, Friesen WV (1975) Unmasking the face. Prentice Hall, Englewood Cliffs
- Elster J (1998) Emotions and economic theory. *J Econ Lit* 36:47–74
- Etcoff NL (1989) Recognition of emotion in patients with unilateral brain damage. In: Gainotti G, Caltagirone C (eds) *Emotions and the dual brain. Experimental brain research series* 18. Springer, New York, pp 168–186
- Ferrer RA, Fisher JD, Buck R, Amico KR (2011) Adding emotional education to a social-cognitive sexual risk reduction intervention may sustain behavior change. *Health Psychol* (ePub ahead of print)
- Fiedler K (2000) Towards an integrative account of affect and social cognition phenomena. In:Forgas JP (ed) *Feeling and thinking: the role of affect in social cognition*. Cambridge University press, Cambridge, pp 223–252
- Fine C, Lumsden J, Blair RJR (2001) Dissociation between 'theory of mind' and executive functions in a patient with early left amygdala damage. *Brain* 124:287–298
- Finkenauer C, Gallucci M, van Dijk WW, Pollmann M (2007) Investigating the role of time in affective forecasting: temporal influences on forecasting accuracy. *Pers Soc Psychol Bull* 33:1152–1166
- Finucane ML (2008) Emotion, affect, and risk communication with older adults: challenges and opportunities. *J Risk Res* 11:983–997
- Finucane ML, Peters E, Slovic P (2003) Judgment and decision making: the dance of affect and reason. In: Schneider SL, Shanteau J (eds) *Emerging perspectives on judgment and decision research*. Cambridge University Press, Cambridge, pp 327–364
- Fischhoff B, Slovic P, Lichtenstein S, Read S, Combs B (1978) How safe is safe enough? A psychometric study of attitudes toward technological risks and benefits. *Policy Sci* 9:127–152
- Fischhoff B, Gonzalez RM, Lerner JS, Small DA (2005) Evolving judgments of terror risks: foresight, hindsight, and emotion. *J Exp Psychol Appl* 11:124–139
- Forgas JP (1995) Mood and judgment: the affect infusion model (AIM). *Psychol Bull* 117:39–66
- Forgas JP (1999) On feeling good and being rude: affective influences on language use and request formulations. *J Pers Soc Psychol* 76:928–939
- Forgas JP, Bower GH, Moylan SJ (1990) Praise or blame? Affective influences in attributions for achievement. *J Pers Soc Psychol* 59:809–818
- Gasper K, Clore GL (1998) The persistent use of negative affect by anxious individuals to estimate risk. *J Pers Soc Psychol* 74:1350–1363
- Gilbert DT, Pinel EC, Wilson TD, Blumberg SJ, Wheatley TP (1998) Immune neglect: a source of durability bias in affective forecasting. *J Pers Soc Psychol* 75:617–638
- Greifender R, Bless H, Kuschmann T (2007) Extending the brand image on new products: the *facilitative effect of happy mood states*. *J Consum Behav* 6:19–31
- Griskevicius V, Goldstein NJ, Mortensen CR, Sundie JM, Cialdini RB, Kenrick DT (2009) Fear and loving in Las Vegas: evolution, emotion, and persuasion. *J Mark Res* 46:384–395
- Griskevicius V, Shiota MN, Neufeld SL (2010) Influence of different positive emotions on persuasion processing: a functional evolutionary approach. *Emotion* 10:190–206
- Grunberg NE, Straub RO (1992) The role of gender and taste class in the effects of stress on eating. *Health Psychol* 11:97–100
- Gur RC, Mozley LH, Mozley DP, Resnick SM, Karp JS, Alavi A, Arnold SE, Gur RE (1995) Sex differences in regional cerebral glucose metabolism during resting state. *Science* 267:528–531
- Han S, Lerner JS, Keltner D (2007) Feelings and consumer decision making: the appraisal-tendency framework. *J Consum Psychol* 17:158–168
- Hardee JE, Thompson JC, Puce A (2008) The left amygdala knows fear: laterality in the amygdala response to fearful eyes. *Soc Cogn Affect Neurosci* 3:47–54
- Harle KM, Sanfey AG (2007) Incidental sadness biases social economic decisions in the ultimatum game. *Emotion* 7:876–881
- Harmon-Jones E, Gable PA, Peterson CK (2010) The role of asymmetric frontal cortical activity in emotion-related phenomena: a review and update. *Biol Psychol* 84:451–462

- Hastings G, Stead M, Webb J (2004) Fear appeals in social marketing: strategic and ethical reasons for concern. *Psychol Mark* 21:961–986
- Hilton D (2008) Emotional tone and argumentation in risk communication. *Judgment Decis Mak* 3:100–110
- Holstege G, Georgiadis JR, Paans AMJ, Meiners LC, Ferdinand HCE, van der Graaf FHCE, Simone Reinders AAT (2003) Brain activation during human male ejaculation. *J Neurosci* 23(27): 9185–9193
- Hsee CK, Weber EU (1997) A fundamental prediction error: self-other discrepancies in risk preference. *J Exp Psychol Gen* 126:45–53
- Janszky J, Szucs A, Halasz P, Borbely C, Hollo A, Barsi P et al (2002) Orgasmic aura originates from the right hemisphere. *Neurology* 58:302–304
- Johnson EJ, Tversky A (1983) Affect, generalization, and the perception of risk. *J Pers Soc Psychol* 45:20–31
- Keltner D, Haidt J, Shiota MN (2006) Social functionalism and the evolution of emotions. In: Schaller M, Simpson JA, Kenrick DT (eds) *Evolution and social psychology*. Psychosocial Press, Madison, pp 115–142
- Killgore WD, Yurgelun-Todd DA (2001) Sex differences in amygdala activation during the perception of facial affect. *Neuroreport* 12:2543–2547
- Killgore WDS, Oki M, Yurgelun-Todd DA (2000) Sex-specific developmental changes in amygdala responses to affective faces. *Neuroreport* 12: 427–433
- Klein WMP, Zajac LE, Monin MM (2009) Worry as a moderator of the association between risk perceptions and quitting intentions in young adult and adult smokers. *Ann Behav Med* 38(3):256–261
- Kluver H, Bucy PC (1939) Preliminary analysis of functions of the temporal lobe in monkeys. *Arch Neurol Psych Chicago* 42:979–1000
- Koenigsberg HW, Sievera LJ, Lee H, Pizzarello S, Newa AS, Goodman M, Cheng H, Flory J, Prohovnik I (2009) Neural correlates of emotion processing in borderline personality disorder. *Psychiatry Res Neuroimaging* 172:192–199
- Kowta S, Buck R (1995) A cross-cultural study of product involvement using the ARI model and the CASC scale. In: Paper presented at the meeting of the American Psychological Association, New York, 12 Aug 1995
- Kunst-Wilson WR, Zajonc RB (1980) Affective discrimination of stimuli that cannot be recognized. *Science* 207:557–558
- Lawton R, Conner M, McEachan R (2009) Desire or reason: predicting health behaviors from affective and cognitive attitudes. *Health Psychol* 28:56–65
- Lazarus RS (1982a) Thoughts on the relations between affect and cognition. *Am Psychol* 37:1019–1024
- Lazarus RS (1982b) On the primacy of cognition. *Am Psychol* 39:124–129
- Lazarus RS (1991a) Emotion and adaptation. Oxford University Press, Oxford
- Lazarus RS (1991b) Cognition and motivation in emotion. *Am Psychol* 46:352–367
- Lazarus RS (1994a) Universal antecedents of emotion. In: Ekman P, Davidson RJ (eds) *The nature of emotion: fundamental questions*. Oxford University Press, New York, pp 163–171
- Lazarus RS (1994b) Appraisal: the long and the short of it. In: Ekman P, Davidson RJ (eds) *The nature of emotion: fundamental questions*. Oxford University Press, New York, pp 208–215
- Lazarus RC, Alfert E (1964) Short-circuiting of threat by experimentally altering cognitive appraisal. *J Abnorm Soc Psychol* 69:195–205
- LeDoux J (1986) The neurobiology of emotion. In: LeDoux J, Hirst W (eds) *Mind and brain: dialogues in cognitive neuroscience*. Cambridge University Press, New York, pp 301–354
- LeDoux J (1996) *The emotional brain*. Simon & Schuster, New York
- LeDoux JE (1994) Cognitive-emotional interactions in the brain. In: Ekman P, Davidson RJ (eds) *The nature of emotion: fundamental questions*. Oxford University Press, New York, pp 216–223
- LeDoux JE, Sakaguchi A, Reis DJ (1984) Subcortical efferent projections of the medial geniculate nucleus mediate emotional responses conditioned to acoustic stimuli. *J Neurosci* 4:683–698
- Lerner JS, Keltner D (2000) Beyond valence: toward a model of emotion-specific influences on judgment and choice. *Cogn Emotion* 14:473–494
- Lerner JS, Keltner D (2001) Fear, anger, and risk. *J Pers Soc Psychol* 81:146–159
- Lerner JS, Gonzalez R, Small D, Fischhoff B (2003) Effects of fear and anger on perceived risk of terrorism: a national field experiment. *Psychol Sci* 14:144–150
- Lerner JS, Small DA, Loewenstein GF (2004) Heart strings and purse strings: carry-over effects of emotions on economic transactions. *Psychol Sci* 15:337–341
- Lipkus IM, Klein WMP, Skinner CS, Rimer BK (2005) Breast cancer risk perceptions and breast cancer worry: what predicts what? *J Risk Res* 8:439–452
- Loewenstein G (1996) Out of control: visceral influences on behavior. *Organ Behav Hum Decis Process* 65:272–292
- Loewenstein GF, Lerner JS (2003) The role of affect in decision making. In: Davidson R, Scherer K, Goldsmith H (eds) *Handbook of affective science*. Oxford University Press, New York, pp 619–642
- Loewenstein GF, Weber EU, Hsee CK, Welch N (2001) Risk as feelings. *Psychol Bull* 127:267–286

- MacLean PD (1993) Cerebral evolution of emotion. In: Lewis M, Haviland J (eds) *Handbook of emotions*. Guilford, New York, pp 67–83
- Markowitz HJ (1998) Differential contribution of right and left amygdala to affective information processing. *Behav Neurol* 11:233–244
- McQuarrie EF, Munson JM (1987) The Zaichkowsky personal involvement inventory: modification and extension. In: *Advances in consumer research*, vol 14. Association for Consumer Research, Ann Arbor, pp 36–40
- McWhirter JJ (1995) Emotional education for university students. *J Coll Stud Psychother* 10:27–38
- Miller S, Diefenbach MA, Kruus L, Ohls L, Hanks G, Bruner D (2001) Psychological and screening profiles of first degree relatives of prostate cancer patients. *J Behav Med* 24:247–258
- Ness RM, Klaas R (1994) Risk perception by patients with anxiety disorders. *J Nerv Ment Dis* 182:466–470
- Norton TR, Bogart LM, Cecil H, Pinkerton SD (2005) Primacy of affect over cognition in determining adult men's condom-use behavior. *J Appl Soc Psychol* 35:2493–2534
- Panksepp J (1994) A proper distinction between cognitive and affective process is essential for neuroscientific progress. In: Ekman P, Davidson RJ (eds) *The nature of emotion: fundamental questions*. Oxford University Press, New York, pp 224–226
- Peters E, Slovic P (2000) The springs of action: affective and analytical information processing in choice. *Pers Soc Psychol Bull* 26:1465–1475
- Phelps EA, O'Connor KJ, Gatenby JC, Gore JC, Grillon C, Davis M (2001) Activation of the left amygdala to a cognitive representation of fear. *Nat Neurosci* 4:437–441
- Pritchard A, Morgan N (1998) 'Mood marketing' – the new destination branding strategy: a case study of 'Wales' brand. *J Vacation Mark* 4:215–229
- Prodan CI, Orbello DM, Testa JA, Ross ED (2001) Hemispheric differences in recognizing upper and lower facial displays of emotion. *Neuropsychiatry Neuropsychol Behav Neurol* 14:206–212
- Raghunathan R, Pham MT (1999) All negative moods are not equal: motivational influences of anxiety and sadness on decision making. *Organ Behav Hum Decis Process* 79:56–77
- Rolls J (1999) The brain and emotion. Oxford University Press, New York
- Ross E (1981) The aposodias: functional-anatomic organization of the affective components of language in the right hemisphere. *Arch Neurol* 38:561–569
- Ross ED (1992) Lateralization of affective prosody in the brain. *Neurology* 42(Suppl 3):411
- Ross ED, Homan R, Buck R (1994) Differential hemispheric lateralization of primary and social emotions. *Neuropsychiatry Neuropsychol Behav Neurol* 7:1–19
- Rottenstreich Y, Hsee CK (2001) Money, kisses, and electric shocks: on the affective psychology of risk. *Psychol Sci* 12:185–190
- Rozin P, Haidt J, McCauley CR (1993) Disgust. In: Lewis M, Haviland J (eds) *Handbook of emotions*. Guilford, New York, pp 575–594
- Schneider F, Grodd W, Weiss U, Klose U, Mayer KR, Nägele T, Gur RC (1997) Functional MRI reveals left amygdala activation during emotion. *Psychiatry Res* 76:75–82
- Schwarz N (2002) Situated cognition and the wisdom of feelings: cognitive tuning. In: Feldman Barrett L, Salovey P (eds) *The wisdom in feelings*. Guilford, New York, pp 144–166
- Schwarz N, Clore GL (1983) Mood, misattribution, and judgments of well-being: informative and directive functions of affective states. *J Pers Soc Psychol* 45:513–523
- Shamay-Tsoory SG, Lavidor M, Aharon-Peretz J (2008) Social learning modulates the lateralization of emotional valence. *Brain Cogn* 67:280–291
- Shaughnessy MF, Shakesby P (1992) Adolescent sexual and emotional intimacy. *Adolescence* 27:475–480
- Silberman EK, Weingartner H (1986) Hemispheric lateralization of functions related to emotion. *Brain Cogn* 5:322–353
- Slovic P (1987) Perception of risk. *Science* 236:280–285
- Slovic P, Finucane ML, Peters E, MacGregor DG (2004) Risk as analysis and risk as feelings: some thoughts about affect, reason, risk, and rationality. *Risk Anal* 24:311–322
- Smith SD, Abou-Khalil B, Zaid DH (2008) Posttraumatic stress disorder in a patient with no left amygdala. *J Abnorm Psychol* 117:479–484
- Spiesman JC, Lazarus RC, Mordkoff AM, Davidson LA (1964) Experimental reduction of stress based on ego-defense theory. *J Abnorm Soc Psychol* 68:367–380
- Straube T, Weisbrod A, Schmidt S, Raschdorf C, Preul C, Mentzel H-J, Miltner WHR (2010) No impairment of recognition and experience of disgust in a patient with a right-hemispheric lesion of the insula and basal ganglia. *Neuropsychologia* 48:1735–1741
- Strauss E (1986) Cerebral representation of emotion. In: Blanck P, Buck R, Rosenthal R (eds) *Nonverbal communication in the clinical context*. Pennsylvania State University Press, University Park/London, pp 176–196
- Szabo CA, Xiong J, Lancaster JL, Rainey L, Fox P (2001) Amygdalar and hippocampal volumetry in control participants. *Am J Neuroradiol* 22: 1342–1345

- Thaler RH, Sustein CR (2008) *Nudge: improving decisions about health, wealth, and happiness*. Yale University Press, New York
- Tucker DM (1981) Lateral brain function, emotion, and conceptualization. *Psychol Bull* 89:19–46
- Weinstein ND (1989) Effects of personal experience on self-protective behavior. *Psychol Bull* 105:31–50
- Weinstein ND, Kwtel A, McCaul KD, Magnan RE, Gerrard M, Gibbons FX (2007) Risk perceptions: assessment and relationship to influenza vaccination. *Health Psychol* 26:146–151
- Wilson WR (1979) Feeling more than we can know: exposure effects without learning. *J Pers Soc Psychol* 37:811–821
- Zajonc RB (1980) Feeling and thinking: preferences need no inference. *Am Psychol* 35:151–175
- Zajonc R (1984) On the primacy of affect. *Am Psychol* 39:117–123



# 28 Cultural Cognition as a Conception of the Cultural Theory of Risk

Dan M. Kahan

Yale University, New Haven, CT, USA

<i>Introduction</i> .....	726
<i>The Cultural Theory of Risk, in Broad Strokes</i> .....	727
<i>Measuring Worldviews</i> .....	729
Dake and His Successors .....	729
Cultural Cognition .....	730
But Is It <i>Cultural Theory</i> ? .....	733
How Many Cultures? .....	734
Where Have All the Fatalists Gone? .....	735
Where Are Cultural Views Located? Institutions Versus Individuals .....	736
Whose Worldviews? Crosscultural Risk Perception .....	737
<i>Mechanisms</i> .....	739
Cultural Identity-Protective Cognition .....	740
Culturally Biased Assimilation of (and Search for) Information and Cultural Polarization .....	742
Cultural Availability .....	746
Cultural Credibility Heuristic .....	749
Cultural-Identity Affirmation .....	752
<i>Further Research: Collective Management of Cultural Bias</i> .....	753

**Abstract:** Cultural cognition is one of a variety of approaches designed to empirically test the “cultural theory of risk” set forth by Mary Douglas and Aaron Wildavsky. The basic premise of cultural theory is that individuals can be expected to form beliefs about societal dangers that reflect and reinforce their commitments to one or another idealized form of social ordering. Among the features of cultural cognition that make it distinctive among conceptions of cultural theory are its approach to measuring individuals’ cultural worldviews; its empirical investigation of the social psychological mechanisms that connect individuals’ risk perceptions to their cultural worldviews; and its practical goal of enabling self-conscious management of popular risk perceptions in the interest of promoting scientifically sound public policies that are congenial to persons of diverse outlooks.

## Introduction

---

This entry examines two related frameworks for the study of popular risk perceptions: the *cultural theory* of risk, associated with the work of Douglas and Wildavsky (1982); and the *cultural cognition* of risk, a focus of recent work by various researchers including myself. I will present the latter as a conception of the former. The motivation for characterizing cultural cognition as a “conception” of cultural theory is twofold: first, to supply an expositional framework for cultural cognition, the concepts and methods of which were formed to empirically test cultural theory; and second, to emphasize that cultural cognition is only one of a variety of competing approaches for interpreting and testing Douglas and Wildavsky’s influential claims about the nature of risk perception.

Indeed, one premise of this entry is that the answer to the question whether cultural cognition supplies a “correct” understanding of cultural theory, if not entirely unimportant, ultimately has no bearing on whether cultural cognition helps to make sense of individual differences in risk perception. Accordingly, whatever objections one might make to cultural cognition in the name of a particular rendering of Douglas and Wildavsky’s theory does not detract from the explanatory and predictive (and ultimately prescriptive) utility of cultural cognition. Of course, if cultural cognition does a better job than other attempts to operationalize the cultural theory of risk, one might reasonably count this feature of it as a reason to prefer it to other constructs that arguably fit better with some set of theory-derived criteria but that do not, as an empirical matter, conform to the phenomena cultural theory is meant to explain.

There are three features of cultural cognition, I submit, that are distinctive among the various conceptions of cultural theory. One is the way in which cultural cognition *measures* cultural worldviews, which are the primary explanatory variable in the Douglas–Wildavsky account of risk perceptions. Another is the attention that cultural cognition gives to the mechanisms – social and psychological – that explain how culture shapes individuals’ beliefs about risk. And the third is the practical objective of cultural cognition to promote collective management of public perceptions of risk and the effect of policies for mitigating them.

The entry elaborates on these points. Part 2 starts with an overview of cultural theory – a very spare one, which will be filled in over the course of later parts but which suffices to set the exposition in motion. Part 3 will take up the measurement of worldviews, contrasting the methods that I and my collaborators in the Cultural Cognition Project use with those used by other scholars who have tried to test cultural theory empirically. Part 4 addresses the distinctive focus of cultural cognition on psychological mechanisms. I’ll identify how this feature of

cultural cognition does admittedly put it in conflict with an important feature of Douglas and Wildavsky's own view of their theory. But I'll also survey some key empirical findings that cultural cognition has generated by adding psychological mechanisms to cultural theory. Part 5 concludes with an assessment of how cultural cognition can guide research aimed at enabling collective management of the role that culture plays in risk perception, a normative objective that might well strike orthodox cultural theorists as puzzling.

## The Cultural Theory of Risk, in Broad Strokes

Cultural theory asserts that individuals should be expected to form perceptions of risk that reflect and reinforce their commitment to one or another “cultural way of life” (Thompson et al. 1990). The theory uses a scheme that characterizes cultural ways of life and supporting worldviews along two cross-cutting dimensions (Fig. 28.1), which Douglas calls “group” and “grid” (Douglas 1970, 1982). A “weak” group way of life inclines people toward an individualistic worldview, highly “competitive” in nature, in which people are expected to “fend for themselves” without collective assistance or interference (Rayner 1992, p. 87). In a “strong” group way of life, in contrast, people “interact frequently in a wide range of activities” in which they “depend on one another” to achieve their ends. This mode of social organization “promotes values of solidarity rather than the competitiveness of weak group” (Rayner 1992, p. 87).

A “high” grid way of life organizes itself through pervasive and stratified “role differentiation” (Gross and Rayner 1985, p. 6). Goods and offices, duties and entitlements, are all “distributed on the basis of explicit public social classifications such as sex, color,... a bureaucratic office, descent in a senior clan or lineage, or point of progression through an age-grade system” (Gross and Rayner 1985, p. 6). It thus conduces to a “hierarchic” worldview

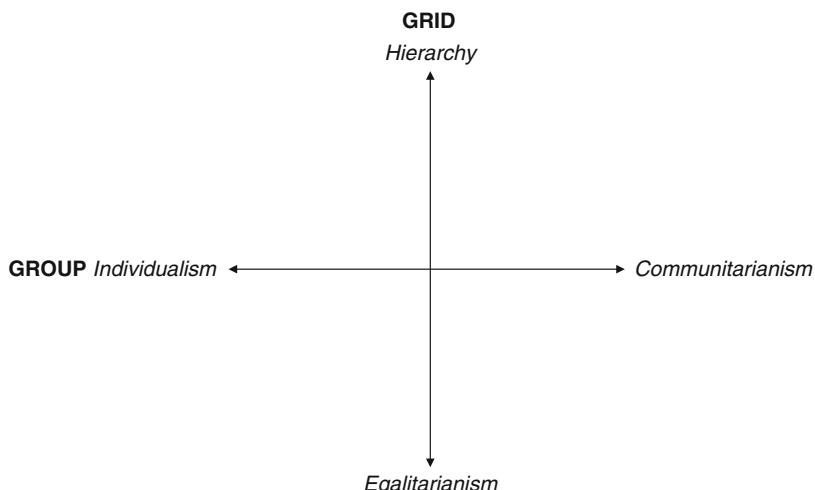


Fig. 28.1

Douglas's “group–grid” scheme. “Group” and “grid” delineate orthogonal dimensions of social organization, or “ways of life,” and supportive values or “worldviews”

that disposes people to “devote a great deal of attention to maintaining” the rank-based “constraints” that underwrite “their own position and interests” (Rayner 1992, p. 87).

Finally, a low grid way of life consists of an “egalitarian state of affairs in which no one is prevented from participation in any social role because he or she is the wrong sex, or is too old, or does not have the right family connections” (Rayner 1992, p. 87). It generates and is supported by a correspondingly egalitarian worldview that emphatically denies that goods and offices, duties and entitlements, should be distributed on the basis of such rankings.

The cultural theory of risk makes two basic claims about the relationship between cultural “ways of life” so defined and risk perceptions. The first is that discrete constellations of perceived risk tend to cohere better with one or another way of life. Forms of conduct understood to inflict collective harm invite restriction, and the people who engage in such behavior censure and blame (Douglas 1992). It thus secures a way of life when its members come to see those who deviate from its norms as exposing the group to risk, in which case “the belief that the innocent are in danger helps to brand the delinquent and to rouse moral fervor against him” (Douglas 1966, p. 134). By the same token, it threatens a way of life, and the authority of those who hold positions of high status within it, to identify its signature forms of behavior as courting collective injury (Douglas and Wildavsky 1982).

The second claim of cultural theory is that individuals gravitate toward perceptions of risk that advance the way of life to which they are committed. “[M]oral concern guides not just response to the risk but the basic faculty of [risk] perception” (Douglas 1985, p. 60). Each way of life and associated worldview “has its own typical risk portfolio,” which “shuts out perception of some dangers and highlights others,” in manners that selectively concentrate censure on activities that subvert its norms and deflect it away from activities integral to sustaining them (Douglas and Wildavsky 1982, pp. 8, 85). Because ways of life dispose their adherents selectively to perceive risks in this fashion, disputes about risk, Douglas and Wildavsky argue, are in essence parts of an “ongoing debate about the ideal society” (Douglas and Wildavsky 1982, p. 36).

The paradigmatic case, for Douglas and Wildavsky, is environmental risk perception. Persons disposed toward the individualistic worldview supportive of a weak group way of life should, on this account, be disposed to react very dismissively to claims of environmental and technological risk because they recognize (how or why exactly is a matter to consider presently) that the crediting of those claims would lead to restrictions on commerce and industry, forms of behavior they like. The same orientation toward environmental risk should be expected for individuals who adhere to the hierarchical worldview, who see assertions of such danger as implicit indictments of the competence and authority of societal elites. Individuals who tend toward the egalitarian and solidaristic worldview characteristic of strong group and low grid, in contrast, dislike commerce and industry, which they see as sources of unjust social disparities, and as symbols of noxious self-seeking. They therefore find it congenital to credit claims that those activities are harmful – a conclusion that does indeed support censure of those who engage in them and restriction of their signature forms of behavior (Wildavsky and Dake 1990; Thompson et al. 1990).

This was the plot of Douglas and Wildavsky’s classic *Risk and Culture: An Essay on the Selection of Technological and Environmental Dangers* (1982). The relationship that Douglas and Wildavsky asserted there between risk perceptions and the various ways of life featured in group-grid has animated nearly two decades’ worth of empirical research aimed at testing the cultural theory of risk.

## Measuring Worldviews

One of central methodological issues in the empirical research inspired by Douglas and Wildavsky's cultural theory is how to measure cultural worldviews. The approach cultural cognition takes toward this task is one of the things that distinguishes it from other conceptions of cultural theory.

### Dake and His Successors

The dominant approach to measuring cultural worldviews can be traced back to Karl Dake, who (along with Wildavsky, his Ph.D. dissertation advisor) published the first empirical studies of culture theory in the early 1990s (Dake 1990, 1991; Wildavsky and Dake 1990). The basis of these studies was a pair of public opinion surveys of residents of San Francisco and Oakland, California. Although the survey instrument was not designed specifically to test cultural theory, Dake was able to use various items from it to construct measures for doing so. Thus, from items relating to respondents' political attitudes, Dake formed separate scales for "Hierarchy," "Egalitarianism," and "Individualism." In subsequent work (Dake 1992), he identified a fourth set of items to represent a "fatalist" worldview, and thereafter identified the four scales with the quadrants demarcated by the intersection of "group" and "grid" (☞ [Fig. 28.2](#)). Among the

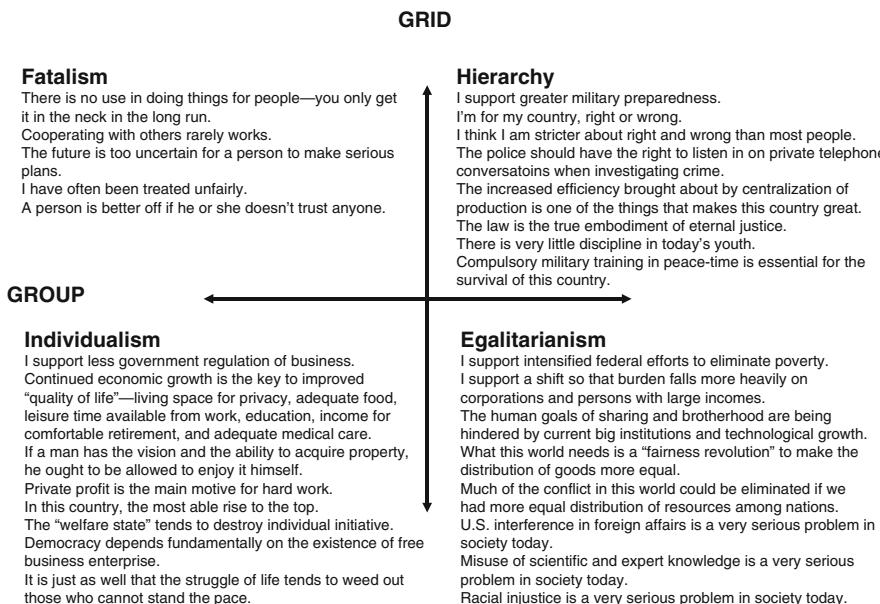


Fig. 28.2

Dake's culture scales. Dake used these items to construct scales for measuring the "worldviews" associated with the group-grid quadrants. The Hierarchy, Egalitarianism, and Individualism scales were used to measure attitudes and analyze risk perceptions in his dissertation (1990) and were the basis of study results published by Dake (1991) and Wildavsky and Dake (1990). The "fatalism" scale was proposed later (Dake 1992)

many studies using Dake's measures or refinements thereof are Ellis and Thompson (1997); O'Connor et al. (1998); Peters and Slovic (1996); Langford et al. (2000); Jenkins-Smith (2001).

Studies based on Dake's measures have encountered two difficulties. The first has to do with the psychometric properties of the various scales. Dake himself did not report any measures of scale reliability. But subsequent researchers have investigated this matter in depth, and they have often found that the separate scales used to measure the respective worldviews perform poorly, failing to display internal validity in tests such as Cronbach's *alpha* (Sjöberg 1998a; Gastil et al. 1995; Marris et al. 1998).

The second problem is conceptual in nature. When one uses separate scales to measure each group–grid worldview, it becomes theoretically possible for a single individual to exhibit multiple, competing orientations – for example, to be simultaneously both a hierarchist and an egalitarian. Indeed, most likely because the items associated with discrete scales do not reflect a high degree of coherence or internal consistency; it's not uncommon for subjects to have high scores on competing scales (Marris et al. 1998). This feature of the Dake scales makes them unsuited for empirically testing cultural theory. Douglas and Wildavsky asserted that individuals attend selectively to risk in patterns that reflect and promote the ways of life to which they subscribe. That claim cannot be cleanly tested with measures that permit individuals to be characterized as subscribing to mutually inconsistent worldviews, for in that case Douglas and Wildavsky's position doesn't yield any determinate predictions about which risks they will credit and which they will dismiss.

## Cultural Cognition

Some researchers, most notably, Hank Jenkins-Smith and his collaborators (Jenkins-Smith and Herron 2009; Silva and Jenkins-Smith 2007; Jenkins-Smith 2001), have made considerable progress in remedying these problems through refinement of Dake's measures. Cultural cognition, however, seeks to avoid them by departing more radically from Dake's strategy for measuring worldviews (► Fig. 28.3).

Cultural cognition uses two continuous attitudinal scales. One, “hierarchy–egalitarianism,” consists of items that determine a person's relative orientation toward high or low “grid” ways of life. The other, “individualism–communitarianism” (we found that readers sometimes were confused by “solidarism,” the term we originally selected to denote the orientation opposed to individualism), consists of items that determine a person's relative orientation toward weak or strong “group” ways of life. In studies performed on general population samples in the USA, these scales have proven highly reliable (e.g., Kahan et al. 2007, 2009, 2010b). They also avoid the logical indeterminacy problem associated with variants of Dake's original measures. When one uses a single scale for group and a single scale for grid, each individual respondent's worldview is identified with a unique point or coordinate in the “culture space” demarcated by the intersection of group and grid (► Fig. 28.4).

More recently, studies using cultural cognition have relied on “short form” versions of the two scales (Kahan et al. 2011). Each short-form scale consists of only six “agree–disagree” items that are “balanced” in attitudinal valance (three items supportive of each end of the two continuous scales). The scales are as reliable as the full-form counterparts. They load appropriately on discrete latent dispositions, generating orthogonal principal components or factors, the scores for which can be used as continuous measures (► Fig. 28.5).

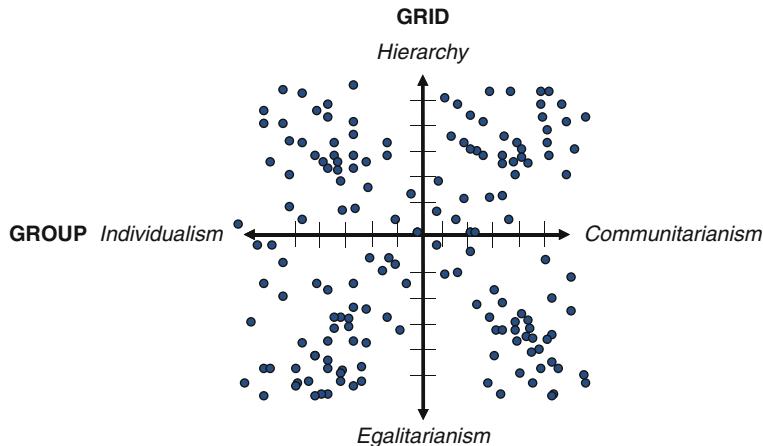
<b>Hierarchy-Egalitarianism</b>		<b>Individualism-Communitarianism</b>	
HCHEATS	It seems like the criminals and welfare cheats get all the breaks, while the average citizen picks up the tab.	IENJOY	People who are successful in business have a right to enjoy their wealth as they see fit.
HEQUAL	We have gone too far in pushing equal rights in this country.	IFIX	If the government spent less time trying to fix everyone's problems, we'd all be a lot better off.
HFEMININ	Society as a whole has become too soft and feminine.	IGOVWAST	Government regulations are almost always a waste of everyone's time and money.
HREVDIS1	Nowadays it seems like there is just as much discrimination against whites as there is against blacks.	IINTRFER	The government interferes far too much in our everyday lives.
HREVDIS2	It seems like blacks, women, homosexuals and other groups don't want equal rights, they want special rights just for them.	IMKT	Free markets—not government programs—are the best way to supply people with the things they need.
HTRADFAM	A lot of problems in our society today come from the decline in the traditional family, where the man works and the woman stays home.	INEEDS	Too many people today expect society to do things for them that they should be doing for themselves.
HWMNRTS	The women's rights movement has gone too far.	INEEDY	It's a mistake to ask society to help every person in need.
EDISCRIM	Discrimination against minorities is still a very serious problem in our society.	IPRIVACY	The government should stop telling people how to live their lives.
EDIVERS	It's old-fashioned and wrong to think that one culture's set of values is better than any other culture's way of seeing the world...	IPROFIT	Private profit is the main motive for hard work.
ERADEQ	We need to dramatically reduce inequalities between the rich and the poor, whites and people of color, and men and women.	IPROTECT	It's not the government's business to try to protect people from themselves.
EROUGH	Parents should encourage young boys to be more sensitive and less "rough and tough."	IRESPON	Society works best when it lets individuals take responsibility for their own lives without telling them what to do.
EWEALTH	Our society would be better off if the distribution of wealth was more equal.	ITRIES	Our government tries to do too many things for too many people. We should just let people take care of themselves.
EXSEXIST	We live in a sexist society that is fundamentally set up to discriminate against women.	CHARM	Sometimes government needs to make laws that keep people from hurting themselves.
		CLIMCHOI	Government should put limits on the choices individuals can make so they don't get in the way of what's good for society.
		CNEEDS	It's society's responsibility to make sure everyone's basic needs are met.
		CPROTECT	The government should do more to advance society's goals, even if that means limiting the freedom and choices of individuals.
		CRELY	People should be able to rely on the government for help when they need it.

Fig. 28.3

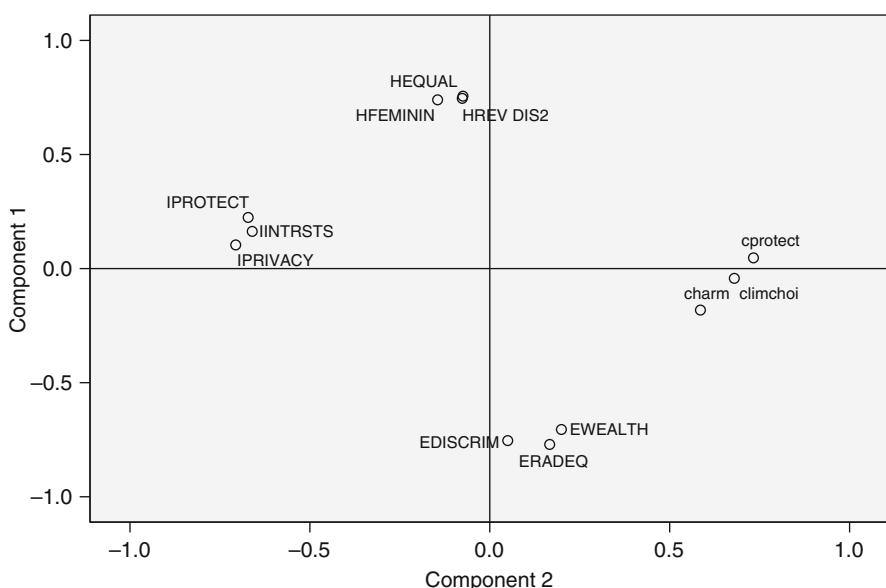
**Cultural cognition scales (“full form”). Study participants indicate the level of their “disagreement” or “agreement” with each item on a four- or six-point Likert response measure. Responses are then aggregated (with appropriate reverse-coding of the “E” and “C” items) to form continuous “Hierarchy-egalitarianism” and “Individualism-communitarianism” worldview scores. When these items are administered to US general population samples, Cronbach’s alpha for each worldview scale consistently exceeds 0.70**

These properties of the scales make them well-suited for testing Douglas and Wildavsky’s theory. If “dangers are culturally selected for recognition” (Douglas 1985, p. 54), then there should be a significant correlation between individuals’ perceptions of risks hypothesized to promote one or another combination of worldviews and the position of individuals’ own worldviews on the “group–grid” map.

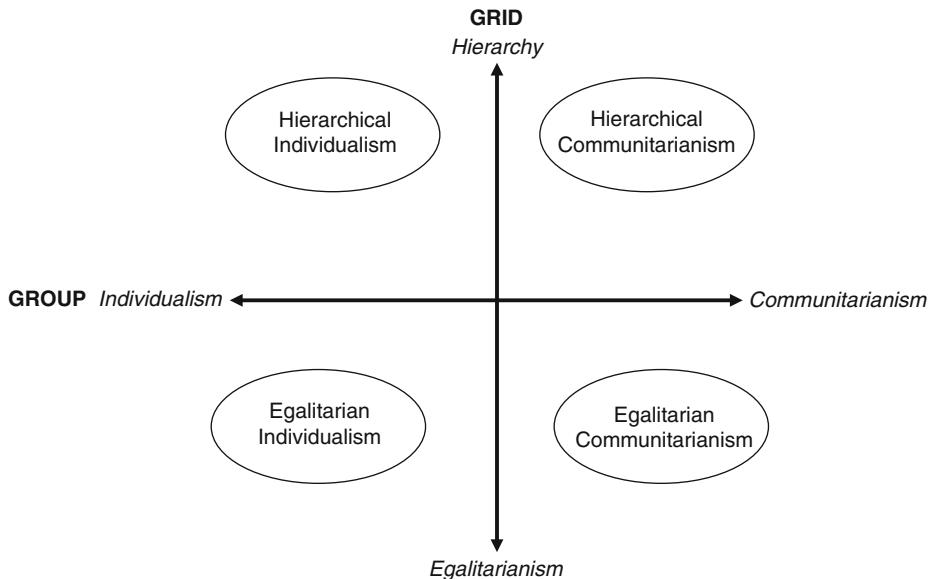
Psychometrically speaking, the scales should be thought of as measures of *latent* or unobserved dispositions, for which the items that make up the scales are simply observable indicators. Because the scales are continuous, they lend themselves readily to correlational analyses (including multivariate regression) in which their influence can be assessed without the loss of statistical power (and the potential bias) associated with splitting a sample into subgroups (Judd 2000).

**Fig. 28.4**

Cultural cognition “map.” The cultural cognition scales can be used to plot the location of any individual on a cultural cognition map (I’d certainly like to say “on a grid” – but you can see how confusing that would become) based on that individual’s scores on the “hierarchy–egalitarianism” and “individualism–communitarianism” scales

**Fig. 28.5**

**Short-form culture scales.** Short forms for Individualism–communitarianism (Cronbach’s  $\alpha = 0.76$ ) and Hierarchy–egalitarianism (Cronbach’s  $\alpha = 0.84$ ), each of which consists of six items loading on orthogonal principal components

**Fig. 28.6**

Cultural cognition “ways of life.” Cultural cognition scales contemplate that combinations of high and low values could have distinctive effects, which admit of measurement through a product interaction term

Cultural theory assigns distinctive effects to the combinations of worldviews supportive of the “ways of life” that inhabit the quadrants of “group-grid.” We designate these four ways of life “hierarchical individualism,” “hierarchical communitarianism,” “egalitarian individualism,” and “egalitarian communitarianism” (● Fig. 28.6), labels we believe are intuitive and appropriately descriptive – albeit different from the diverse array of labels that cultural theorists tend to use (a matter I will return to presently). Even when group and grid are conceptualized as two orthogonal continuous worldview dimensions, the use of an interaction term will make it possible to take account of any unique effects associated with particular combinations of low and high values on the two scales. Added, say, to a multivariate regression, such a term reports the impact of each worldview dimension conditional on a person’s location on the other (see, generally, Aiken et al. 1991). As a result, the effect of a “hierarchical individualist” worldview, say, can have an effect different from (perhaps larger, perhaps smaller than) the one derived by simply adding a low group score and a high grid one.

### But Is It *Cultural Theory*?

However successful it might be in making “cultural worldviews” psychometrically tractable, this approach arguably faces problems of its own related to its fit with cultural theory. I will identify some of these, the elaboration of which also helps to paint a more vivid picture of cultural theory and its intricacies.

## How Many Cultures?

One difficulty has to do with what Thompson et al. (1990) refer to as the “impossibility theorem.” The “impossibility theorem” posits that there are a finite number of viable ways of life – five, according to Thompson et al. (1990), four according to many other cultural theorists – within the space demarcated by group–grid (☞ Fig. 28.7). Because cultural cognition measures treat group and grid as continuous, it might be understood to imply that there can be an infinite number of ways of life formed by congregations of persons around any coordinate in the group–grid map. The impossibility theorem says that’s “impossible” – only five (or maybe just four) ways of life are viable.

To this point – which has been made to us by various cultural theorists – I myself would say the appearance of tension between the cultural cognition scheme and the impossibility theorem might well be illusory. Even if only a limited number of ways of life exist or are “possible,” it doesn’t follow, logically, that every individual must display a worldview that perfectly maps onto these ways of life. A certain measure of heterogeneity among individuals is perfectly consistent with there being aggregations of persons who exert a dominant influence on social structures and affiliated worldviews (Braman et al. 2005). Under either of these conditions, we would expect individuals to form packages of risk perceptions characteristic of their groups in proportion to the strength or degree of attachment to the cultural groups with

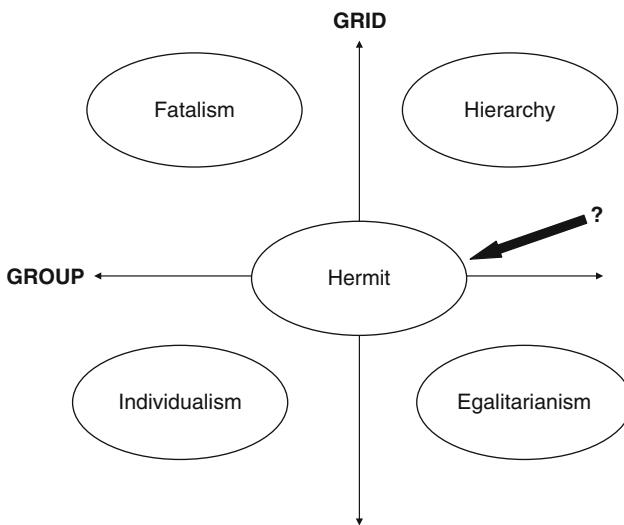


Fig. 28.7

“Impossibility theorem” – only 5 (or maybe only 4). Certain cultural theorists asserts that the number of viable, discrete ways of life (and associated worldviews) within the group–grid scheme is finite, a position known as the “impossibility theorem,” which refers to the impossibility of intermediate ways of life in the interstices of the identified ones. Thompson et al. (1990) posit five such ways of life. The “existence and position on the map” of the “Hermit” are “much disputed,” it is reported, in debates internal to cultural theory (Mamadouh 1999, p. 401)

whom they are most closely affiliated (cf. Manton et al. 1992). That's basically what our measures are designed to show.

But in any case, no one, to my knowledge, has ever purported to empirically test, much less vindicate, the "impossibility theorem," and at least some cultural theorists do indeed take the position (including Mary Douglas at certain points but not at others) that group and grid are inherently spectral in nature and capable of supporting any number of different coherent ways of life within the space demarcated by their intersection (Mamadouh 1999). My collaborators and I take a pragmatic attitude: we are more interested in finding a scheme for measuring cultural worldviews that is internally valid and that has explanatory utility than in finding one that fits a profile dictated by axiomatic, abstract theorizing.

## Where Have All the Fatalists Gone?

Another theory-based objection to the cultural cognition scheme is that it ignores "fatalism." Fatalism is a way of life that Douglas and many other culture theorists associated with weak group, high grid. It is said to generate a worldview that disposes people to accept the diminishment of personal agency and a corresponding perception that steps to abate risk will be futile (Thompson et al. 1990).

The constructs measured by the cultural cognition scales do not imply that weak group, high grid social relations will result in a fatalistic way of life or worldview. Rather, this combination of dispositions will cohere with modes of life in which people, as individualists, are strongly resistant to regulation of affairs by remote, collectivist-minded authorities, but still organize their local institutions in highly regimented, and highly stratified, ways. Think of the iconic American cowboy, the "Marlboro Man": He bridles at outside interference with the operation of his ranch, yet still exerts authority over a small community whose members – from ranch hands, to wives, to sons and daughters – all occupy scripted, hierarchical social roles. He is likely to form a dismissive attitude toward environmental risks (the contribution to global warming associated with his cows' methane emissions, say). However, he might be very concerned that various forms of social deviance could threaten order and generate other bad collective consequences (the residents of Broke Back Mountain, on his view, are destined for calamity). He is no fatalist – he has lots of agency, and he is selectively, not uniformly, risk sensitive.

My own response to this disconnect between the cultural cognition scales and conventional culture theory is a sort of shrug of the shoulders. In truth, I've never gotten the theoretical explanation of why weak group, high grid would generate a "fatalistic" way of life. Indeed, I have a hard time even understanding how fatalism could be a *group way of life*, or why a fatalist stance toward risk could be identified as a worldview as opposed to, say, a personality trait of some kind (possibly one the baleful effects of which could be treated with appropriate pharmaceutical interventions). I think the way of life of the Marlboro Man supplies a more cogent account of what one expects to see with the convergence of hierarchy and individualism. It also happens to generate testable predictions, ones our research confirms, about the distinctive schedule of risk perceptions that people with those worldviews are likely to form. Therefore, I'm inclined to give the infamous "so?" response of Dick Cheney – a Marlboro Man if ever there was one – to the complaint that cultural cognition disregards fatalism.

But I would add to this response that other conceptions of cultural theory also display ambivalence toward fatalism. Dake's original work did not include a fatalism scale, and his analyses reporting the results were confined to assessments of the risk-shaping impact of Hierarchy, Egalitarianism, and Individualism (Wildavsky and Dake 1990; Dake 1990, 1991). In addition, Wildavsky, in his individual writings on how culture influences risk perception and mass political opinion, always left fatalism out of the story too (e.g., Wildavsky 1991), maybe because he was likewise puzzled by it.

## Where Are Cultural Views Located? Institutions Versus Individuals

Yet another conflict between cultural cognition position and at least one understanding of how cultural theory works relates to who should be regarded as the *subject* of cultural worldviews. Cultural cognition theory assumes that cultural worldviews are latent predispositions of individuals (i.e., shared but unobserved orientations that one can measure, with varying degrees of precision, by observable indicators, primarily in the form of professed attitudes). Another view, put forth (at least intermittently) by Stephen Rayner and Michael Thompson, asserts that in fact it is a mistake to see cultural worldviews as fixed or stable features of individuals; rather they are immanent properties of institutions, characterized by one or another mode of social organization, that systematically endow individuals with outlooks conducive to the operation of those institutions during the time (but only then) that individuals happen to occupy roles within them (Rayner 1992).

This picture of how culture shapes risk perceptions is at odds not only with cultural cognition. If it is right, then *any* conception of cultural theory that tries to use individual-level measures of worldviews as an explanatory variable is flawed. Individuals are only temporary receptacles of institution-supporting worldviews that get poured into them as they move from place to place. They will thus "flit like butterflies from context to context, chanting the nature of their arguments as they do so" (Rayner 1992, pp. 107–08).

Some might wonder if such a view could ever generate testable predictions. I think it likely could. All we need (and I don't say this facetiously) is a valid and reliable way for measuring what worldviews "contexts" (nuclear power plants; universities; stock trading floors) of different sorts have. We could then randomly move individuals around from one to one, and see if their risk perceptions – on climate change, say, or on gun possession, or nanotechnology – changed in the ways that cultural theory predicts. (We'd need to randomly manipulate the respondents' locations, rather than just going to those places and polling the people we find at them, for as Rayner notes (1992, p. 107), the view that cultural worldviews belong to individuals predicts that they will self-select into institutions that are congenial to their preferred mode of social ordering.) I seriously doubt that such an experiment would confirm the "social mobility hypothesis," as Rayner dubs it. But I do indeed hope that my doubt spurs him and others to conduct the sort of internally valid test that would be required to settle the issue and that to my knowledge no one has yet attempted.

If they succeed, of course, they might or might not have established a conception of cultural theory, but they will definitely have *demolished* the conception that Douglas and Wildavsky appeared to subscribe to, which was obsessed (for understandable reasons) with political conflict over risk. The egalitarians, individualists, and hierarchs in *Risk and Culture* (1982) were groups of people – not buildings or even "sectors" of some sort of multifaceted

system – who disagreed with each other about the ideal society. The gripping and compelling story Douglas and Wildavsky told about such disagreements would have fallen apart if we imagined that the people who were fighting each other on nuclear power or air pollution were just the ones who at any given moment of the day happened to be at home, at work, at the university, at the salon, etc. Similarly, when we try to make sense of “climate change skeptics” and “climate change believers” today – something even Rayner and Thompson would like to do (e.g., Verweij et al. 2006) – we are trying to understand *people* with relatively stable beliefs, not just temporary receptacles for risk outlooks that get poured into them as they wander from place to place. If we want a theory that explains who believes what and why about politically contested empirical claims about risk, then the “social mobility hypothesis” conception of cultural theory is of no use. Or more accurately, if it can be shown that risk perceptions aren’t persistent but vary in individuals as they move from place to place, we should stop trying to do what cultural cognition and all the other conceptions of cultural theory that reflect what Rayner calls the “stability hypothesis” (which he attributes to Douglas herself) try to do – viz., find explanations for patterns of variance in risk perceptions of different groups of people. Because if the “social mobility” position is right, the explanandum of all such conceptions of cultural theory turns out to be just an (very amazing!) illusion.

But there’s simply no point arguing about which – the “mobility” or “stability” hypothesis – is the “right” view of cultural theory in the abstract. The only version of cultural theory that anyone could have any reason to prefer is the one that actually explains the world we live in. So let the matter be resolved by empirical testing.

## Whose Worldviews? Crosscultural Risk Perception

Another objection to cultural cognition as a conception of cultural theory is that it is parochial. We devised our cultural worldview measures because we wanted to understand variance in perceptions of risk within the US public, and didn’t think that Dake’s measures and variants thereof had psychometric properties necessary to allow us to do that. Accordingly, we developed the measures using US subjects, ones who we interviewed in subjects groups and to whom we administered successive versions of our measures, first in writing and then in phone interviews, over a period of years. It has been pointed out to us, by Mary Douglas herself among others (Douglas 2003), that our measures have a distinctly “American feel,” particularly in relation to their picture of hierarchy, which reflects elements of social stratification (e.g., racial ones) that have played a conspicuous part in animating hierarchic modes of social organization historically in the USA. Some critics of Dake’s measures dismissed them on the ground that if you simply translated them into another language – say, Swedish or Portuguese – they did not furnish reliable measures or predict risk perceptions in the manner that cultural theory says they should (Sjöberg 1998b). I suspect our measures, if subjected to the same test, would also perform poorly, even though they work well for US samples.

But, in my view, that test reflects a very odd expectation of what a successful conception of cultural theory should be able to deliver. Douglas, in a position that was very controversial in its own right, did indeed suggest that the group-grid framework would have an element of universality to it, supplying worldview constructs that could be used to make sense of conflicts over risk across place and time. Let’s grant that she was right (but only for the sake of moving forward; it is a bold claim that merits testing, not an axiom to be dogmatically asserted or

recklessly assumed). It is another matter entirely to say that the *indicators* of the latent dispositions associated with these worldviews must be the same everywhere and forever. Why would we think that when we ask a Hadzabe bushman – or even a Swede or Brazilian – would react the way a contemporary American does to the proposition that “it seems like the criminals and welfare cheats get all the breaks, while the average citizen picks up the tab?” What reason is there to think the two will attach the same *meaning* to this proposition (or that the former will even attach any to it)? If they don’t, then this item won’t be a valid or reliable measure of any sort of latent disposition they happen to share. Because the theory is that differences in a latent characteristic explain variation in risk perception, the way to test the theory is to develop observable indicators that are reliable and valid for that latent characteristic *in* the sample one is studying.

I’d also say that while it’s plausible that the same cultural predispositions toward risk will help to explain variance everywhere, it isn’t necessarily going to be the case that the variance they explain is the same in all places. Douglas teaches that risks grab individual attention and become the currency of blame because of what risk-taking behavior connotes about the authority and legitimacy of contested social orderings: “Each culture must have its notions of dirt and defilement which are contrasted with its notion of the positive structure which must not be negated” (Douglas 1966, p. 160). Those connotations, she recognized full well, will also be a matter of decisive historical contingency. The meanings that made ancient Jews believe that defiance of the commandments of Yahweh would cause him to “strike [them] with consumption, and with fever and with inflammation and with fiery heat and with the sword and with blight and with mildew” (*Deuteronomy* 28:22 (1997); quoted in Douglas 1966, p. 51) were unique to them. If we found today at similar coordinates in the group–grid map a group in the Upper West Side of Manhattan, we would not expect them to attribute floods and fires to the profaning of God – but we *might* expect them to attribute those *very* things to forms of behavior (corporate industrialization and excessive personal consumption) that bear meanings that defy *their* shared commitments. So by the same reasoning, why should hierarchy today dictate the same posture toward the risk of carrying guns in societies of historical experiences as diverse as, say, those of England and the USA? In other words, the validity – and value – of a theory that predicts *that* individuals of opposing predispositions will mobilize themselves into opposing factions over risk doesn’t depend on it being able to say, in a manner oblivious to the historical circumstances of such people, *what* that dispute will be about (Kahan and Braman 2003b).

Recently, social psychology has begun to explore “crosscultural” differences in cognition generally. This line of work focuses on identifying society-level variation, typically between members of “Eastern” (generally, Asian) and “Western” (European) nations. Differences in how members of such societies individuate collective entities (e.g., “schools of fish”) from their individual constituents (“individual fish”) is thought to reflect and reinforce diverse understandings of individual and collective responsibilities and prerogatives (Nisbett 2003). This body of work – which is as fascinating as it is important – is a separate line of inquiry from the one associated with the cultural cognition of risk. It’s possible they might at some point be shown to be connected in some way, although the two seem to reflect different assumptions about the scale of “cultural variation”: whereas the “crosscultural cognition” paradigm envisions that differences in values will manifest themselves at the societal level, the “cultural cognition of risk” focuses on how differences in values result in *intraspacial* conflicts, ones that are likely to be distinctive of conditions that are relatively local in time and space.

## Mechanisms

---

Now I turn to the *mechanisms* of cultural cognition. The idea of “mechanisms of cultural cognition” is meant to be an answer the question *why* individuals are disposed, as Douglas and Wildavsky maintained, to form risk perceptions that cohere with the ways of life they subscribe to.

Douglas and Wildavsky, in separate writings, developed an admittedly *functionalist* answer to this question (Douglas 1986; Thompson et al. 1990). That is, they both took the position that individuals form risk perceptions congenial to their ways of life precisely *because* holding those beliefs about risk cohere with and promote their ways of life. This sort of reasoning – which is associated with classical sociological accounts of ideology – has developed a bad reputation in contemporary social science, which sees it as implausibly attributing agency to collective entities (Boudon 1998). Douglas and Wildavsky were fully aware of this and related objections, and developed ingenious arguments to try to deal with them.

I and the other researchers doing work with cultural cognition take a different tack. The *mechanisms* hypothesis is that worldviews yield risk perceptions through a set of social and psychological processes. The processes are well established; they are the heart of the “*psychometric paradigm*” or psychometric theory of risk pioneered by my collaborator Paul Slovic (2000). What hasn’t been fully recognized until now, our research suggests, is how these social and psychological processes interact with cultural ways of life, generating individual differences in risk perception between people who subscribe to competing worldviews (Kahan et al. 2010a). But importantly, this is not a functionalist account because the social and psychological processes associated with the psychometric paradigm, although different from the ones stressed by rational choice economics, don’t treat the needs of collective entities as the causes of individual behavior but instead derive collective behavior from the interaction of individuals self-consciously pursuing fulfillment of their own ends (Balkin 1998; Elster 1985).

Although this marriage of cultural theory and the psychometric theory of risk wasn’t one that Mary Douglas herself ever sanctioned, she at least recognized such a union as a possible strategy for showing how cultural theory works. She did this in a famous essay, “The Depoliticization of Risk” (1997), which specifically criticized Slovic for failing to explore the interaction of culture and the mechanisms of the psychometric paradigm – a feature of Slovic’s research, she maintained, made it innocent of political conflict over risk. “If we were invited to make a coalition between group-grid theory and psychometrics,” Douglas wrote, “it would be like going to heaven” (Douglas 1997, p. 132). In a sense, cultural cognition, to which Slovic himself has made major contributions, *is* such an invitation. But Douglas, I own, might not have intended to be taken seriously when she made this remark, and few of the scholars who are most interested in her work today have shown any interest in this strategy for exploiting the full richness of her and Wildavsky’s thoughts on risk perception.

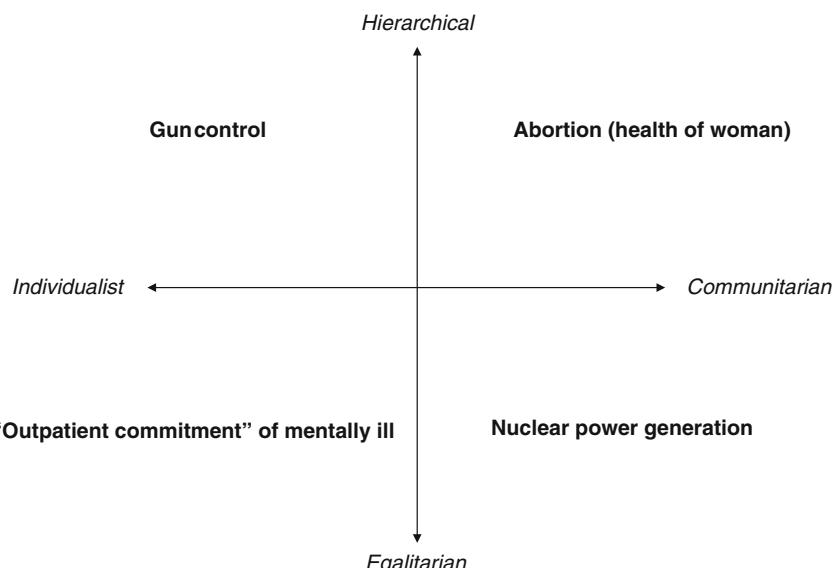
Whether or not viewed as faithful to Douglas’s vision, studies that use psychometric concepts and methods have identified a variety of discrete mechanisms of cultural cognition. The ones I’ll discuss are *identity-protective cognition*; *biased assimilation and group polarization*; *cultural credibility*; *cultural availability*; and *cultural identity affirmation*. I’ll go through them in order, saying something about how each one works in general and then something about studies we’ve done that suggest their influence in connecting cultural worldviews to risk perceptions. This is not an exhaustive list; research is ongoing to investigate additional ones. But these are ones for which there is the best evidence so far.

## Cultural Identity-Protective Cognition

Identity-protective cognition refers to the tendency of people to fit their views to those of others with whom they share some important, self-identifying commitments. Group membership supplies individuals not only with material benefits but a range of important nonmaterial ones, including opportunities to acquire status and self-esteem. Forming beliefs at odds with those held by members of an identity-defining group can thus undermine a person's well-being – either by threatening to drive a wedge between that person and other group members, by interfering with important practices within the group, or by impugning the social competence (and thus the esteem-conferring capacity) of a group generally. Accordingly, individuals are motivated, unconsciously, to conform all manner of attitudes, including factual beliefs, to ones that are dominant within their self-defining reference groups (Cohen 2003; Giner-Sorolla and Chaiken 1997).

The cultural theory of risk holds that groups defined by diverse worldviews can be expected to disagree about risk. Identity-protective cognition furnishes a plausible explanation for why this would be so. One test – of cultural theory generally, and of this particular mechanism for it – would be to determine whether risk perceptions are indeed distributed across groups in patterns that are best explained by the stake individuals have in maintaining the status of, and their status within, groups defined by shared worldviews.

Using our culture scales, we have gathered evidence of such a relationship for a wide variety of risks (Fig. 28.8). Thus, we have been able to show that perceptions of environmental and



**Fig. 28.8**

The distribution of risk perception and cultural identity-protective cognition. Survey evidence establishes that risks associated with the indicated activities are highly correlated with the indicated combination of cultural worldview values even after controlling for other influences. These and like patterns furnish evidence that cultural identity-protective cognition affects the formation of risk perceptions

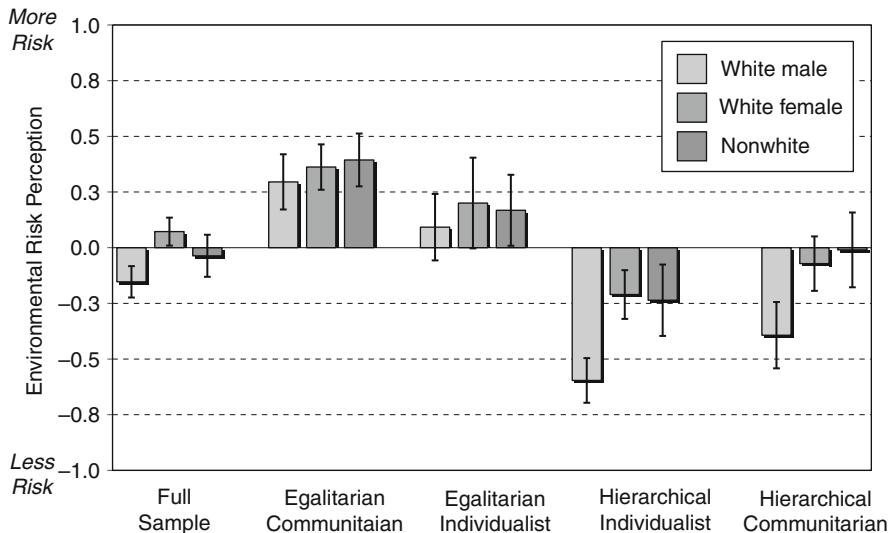
technological risks vary sharply along the lines that Douglas and Wildavsky suggests: that is, as individuals become simultaneously egalitarian and communitarian in their values, they become more concerned, and as they become hierarchical and individualistic less, with climate change, nuclear waste disposal, air pollution, and the like (Kahan et al. 2007). We have also formed and tested our own hypotheses about the distribution of various other risk perceptions that we expected – based largely on ethnographic, historical, and other forms of inquiry – would pit individuals located in one or another quadrant of the group–grid map against those from another. On whether private gun ownership increases or decreases violent crime; on whether abortion impairs the health of women (Kahan et al. 2007); on whether legally compelled submission to medical treatment (including psychotropic drugs) promotes the well-being of mentally ill individuals and the safety of their communities (Kahan et al. 2010a) – in all these cases, we have found that cultural worldviews, as measured with our scales, explained variation better than other individual characteristics, including education, income, personality type, and ideology.

Indeed, the strongest evidence for cultural identity-protective cognition comes from the power of cultural worldviews to explain gender and racial variance in risk perceptions. The “white male effect” refers to the tendency of white males to regard all manner of societal risk as smaller in magnitude and seriousness than do women and minorities (Finucane et al. 2000). We hypothesized (Kahan et al. 2007) that culturally grounded, identity-protective cognition might explain this phenomenon. White males who subscribe to ways of life that feature race and gender differentiation in social roles, our reasoning went, have a special stake in putatively dangerous activities essential to their cultural roles. Accordingly, they should be more powerfully impelled by identity-protective cognition than anyone else to resist the claim that those activities are hazardous for society and should be restricted.

Consider environmental risk perceptions. Hierarchs are disposed to dismiss claims of environmental risks because those claims implicitly cast blame on societal elites. But white male hierarchs, who acquire status within their way of life by occupying positions of authority within industry and the government, have even more of a stake in resisting these risk claims than do hierarchical women, who acquire status mainly by mastering domestic roles, such as mother and homemaker. In addition, *white* hierarchical males are likely to display this effect in the most dramatic fashion because of the correlation between being nonwhite and being an egalitarian.

In a study involving a nationally representative sample of 1,800 US residents, we found strong evidence in support of these hypotheses (Kahan et al. 2007). We found that the race effect in environmental risk perceptions – which persisted even when characteristics such as income, education, and liberal–conservative ideology were controlled for – disappeared once hierarchy and individualism were taken into account. In addition, we found the hypothesized interaction between gender and hierarchy; that is, a disposition toward hierarchy exerted a much stronger effect toward environmental-risk skepticism in men than in women. Indeed, once the extreme risk skepticism of white hierarchical males was taken into account, the gender effect in environmental risk perceptions also disappeared.

We found a similar effect with gun risk perceptions. In the USA, guns enable largely hierarchical roles such as father, protector, and provider, and symbolize hierarchical virtues such as honor and courage. Within hierarchical ways of life, moreover, these are roles and virtues distinctive of men, not women, who again occupy roles that don’t feature gun use. These roles and virtues are also largely associated with being a *white* male, in large part because

**Fig. 28.9**

**Cultural identity-protective cognition and the “white male effect.”** Bars indicate z-score on composite “Environmental risk perception” measure (climate change, air pollution, and nuclear power; Cronbach’s  $\alpha = 0.72$ ). Scores are derived from multivariate regression that included cultural worldview measures, race and gender, and appropriate interaction terms, and controlled for numerous other individual characteristics including education, income, personality type, and political ideology. CIs reflect 0.95 level of confidence. The analysis shows that the sample-wide differential between white males and others is attributable entirely to the extreme risk skepticism of hierarchical white males. The differential is largest among hierarchical individualists. Based on data in Kahan et al. (2007), Table 2, Model 3

of the historical association of guns with maintenance of racial hierarchy in the South. On this account, then, we should expect white hierarchical males to be much more invested in gun possession, and thus to be impelled much more forcefully by identity-protective cognition to resist the claim that guns are dangerous and that gun ownership should be restricted. And this is again exactly what we did find in our national study (● Fig. 28.9).

### Culturally Biased Assimilation of (and Search for) Information and Cultural Polarization

“Biased assimilation and polarization” is a dynamic that characterizes information processing. When individuals are unconsciously motivated to persist in their beliefs, they selectively attend to evidence and arguments, crediting those that reinforce their beliefs and dismissing as noncredible those that contravene them. As a result of this “biased assimilation,” individuals tend to harden in their views when exposed to a portfolio of arguments that variously support and challenge their views. By the same token, when *groups* of individuals who are motivated to persist in opposing beliefs are exposed to balanced information, they don’t converge in their views; as a result of biased assimilation they *polarize* (Lord et al. 1979).

My collaborators and I have hypothesized that identity-protective cognition would unconsciously motivate individuals to assimilate risk information in support of culturally congenial results, and hence drive people with opposing worldviews apart as they consider information (Kahan et al. 2006). If so, *culturally biased assimilation* and polarization could be treated as another mechanism for the sorts of relationships between worldview and risk perception posited by cultural theory.

One study we did to test this possibility focused on nanotechnology risks (Kahan et al. 2009). Nanotechnology involves the creation and manipulation of extremely small materials, on the scale of atoms or molecules, which behave in ways very different from larger versions of the same materials. It's a novel science: about 80% of the American public say they have either never heard of it, or have heard only a little. We did an experiment in which we compared the nanotechnology risk perceptions of subjects to whom we supplied balanced information risk-benefit information to subjects to whom we supplied no information.

The results (Fig. 28.10) confirmed our hypothesis. In the no-information condition, individuals of opposing cultural worldviews held relatively uniform risk perceptions. That's not surprising, since the vast majority of them had never heard of nanotechnology. In the information condition, however, hierarchs and egalitarians, and individualists and communitarians, all formed opposing views. In other words, individuals holding these worldviews attended to the balanced information on nanotechnology in a selective fashion that reinforced their cultural predispositions toward environmental and technological risks generally. As a result, they polarized.

This result thus uses an established mechanism of social psychology – biased assimilation – to ground culturally grounded individual differences in risk perception. But in a reciprocal fashion, it also contributes something a bit back to general understandings of that very mechanism.

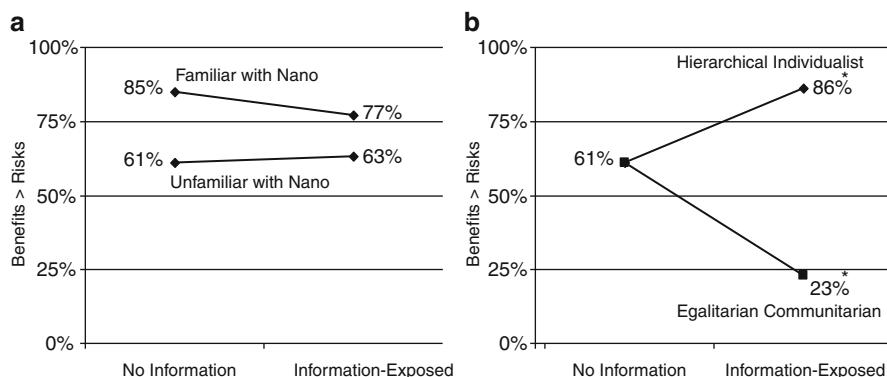


Fig. 28.10

**Biased assimilation/polarization of nanotechnology information.**  $N = 1,850$ . Results derived from statistical simulation based on multivariate logistic regression involving worldviews, information exposure, and prior familiarity with nanotechnology. \* denotes difference between conditions significant at  $p < 0.05$ . Panel (a) shows reactions of subjects to balanced information conditional on prior familiarity and controlling for cultural worldviews. Panel (b) shows reactions of subjects to balanced information conditional on cultural worldviews and controlling for prior familiarity (Kahan et al. 2009)

Conventionally, biased assimilation is determined with reference to individuals' prior beliefs. For example, people who disagree about whether the death penalty deters murder will, when shown studies that reach opposing conclusions, grow even more divided (Lord et al. 1979)

But in our study, individuals, by and large, *had* no priors; most of them said they had not heard anything or anything of significance about nanotechnology before the study. They attended to information, then, in a biased manner supportive of a *predisposition* toward risk. This is a refinement and extension of the biased assimilation/polarization concept. It also attests to the utility of cultural cognition as a *predictive* tool that might be used to anticipate and even, as I will discuss presently, *manage* how diverse persons will react to information about an emerging technology.

Because some of our subjects had in fact learned something about nanotechnology even before the study, we also compared how individuals reacted to information conditional on the level of their prior familiarity. Previous public opinion studies had consistently shown that although relatively few people are aware of nanotechnology, those who are have an extremely positive view of its benefits relative to its risks. This finding has prompted some to infer that as word of nanotechnology spreads, members of the public generally will form positive impressions of this novel science. This is, obviously, a hypothesis at odds with our own.

We found no support for it. It was the case that individuals in our "no information" condition who reported knowing more about nanotechnology had a more favorable view than

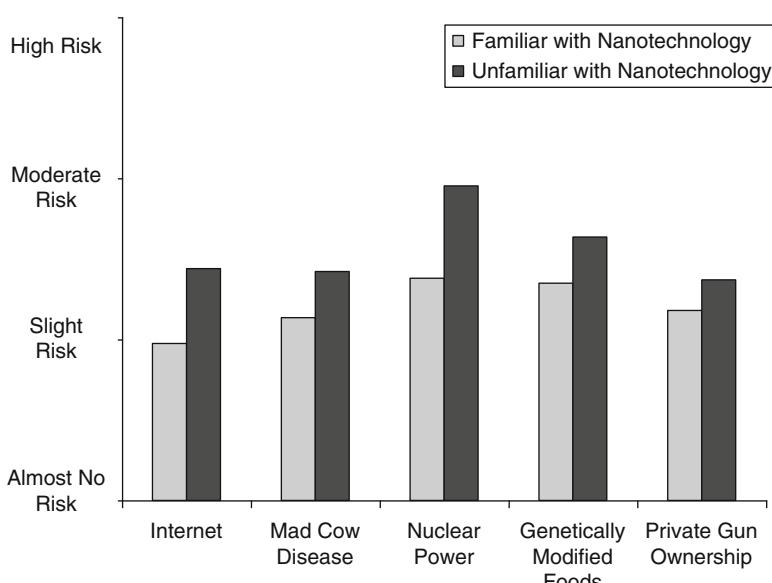


Fig. 28.11

Other risk perceptions among subjects familiar and unfamiliar with nanotechnology.  $n = 1,820$  to  $1,830$ . Risk variables are 4-pt measures of "risk to people in American Society" posed by indicated risk. Differences between group means all significant at  $p \leq 0.01$  (Kahan et al. 2009)

those who reporting knowing nothing or only a little. But when exposed to information, subjects of the latter description did not react in a uniformly positive way; again, they reacted positively or negatively conditional to their cultural worldviews (☞ Fig. 28.10).

We also found out something else interesting about the individuals who claimed they knew a lot or a substantial amount about nanotechnology: they weren't (on average) afraid of anything. They rated the risks of nuclear power as low. They didn't worry about "mad cow disease" or genetically modified foods. They saw owning a gun as low risk too. And so on (☞ Fig. 28.11).

What to make of this? Should the National Rifle Association or the Nuclear Power Chamber of Commerce flood the streams of public communication with information about nanotechnology? Obviously not. This is the signature of spurious correlation: information about nanotechnology is not causing individuals to see guns, the Internet, genetically modified foods, nuclear power, and so forth as safe; some third influence is causing people *both* to form the view that these risks are low *and* to become interested enough in nanotechnology to learn about it before others do.

Why, then, should we not suspect the relationship between familiarity with nanotechnology, on the one hand, and a positive view of its risks and benefits, on the other, as spurious too? Indeed, we'd shown in our experiment that information exposure *doesn't* cause individuals to

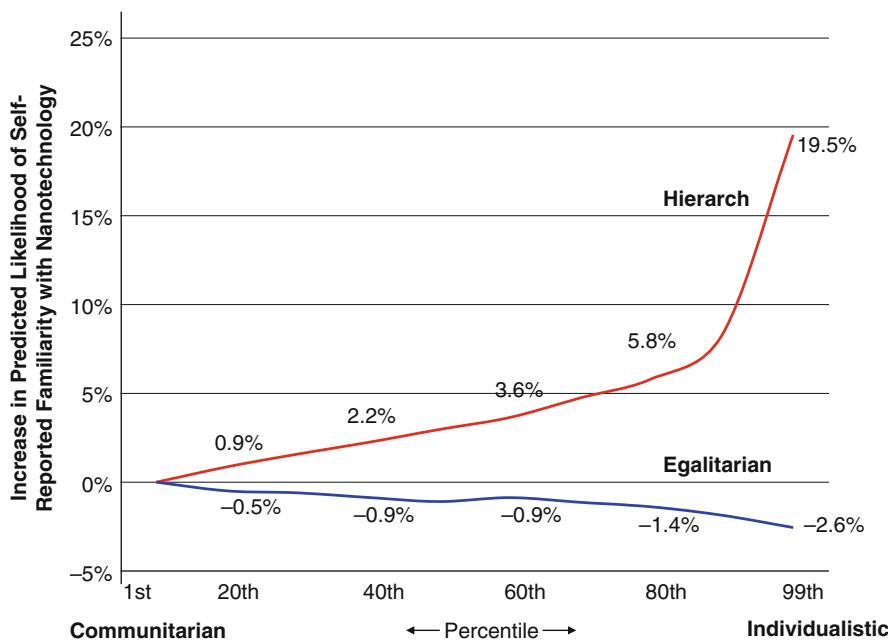
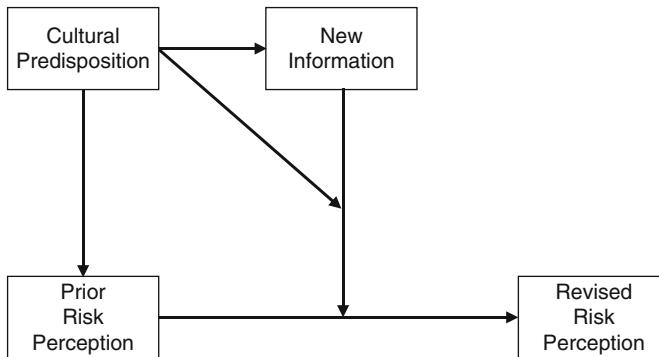


Fig. 28.12

**Predicting familiarity with nanotechnology.**  $N = 1,800$ . Derived from simulation based on multivariate logistic regression with race, sex, gender, education, income, political ideology, and cultural worldview predictors. Interaction between hierarchy and individualism is significant at  $p < 0.05$  (Kahan et al. 2009)

**Fig. 28.13**

**Biased assimilation and search.** Experiment results show that same cultural predisposition responsible for beliefs about environmental risk influence both the search and the effect given to new information

form a positive view. So we tried to see if we could identify a third influence that both causes individuals to learn more about nanotechnology and to see it as low risk.

It shouldn't be hard at this point to guess what we found. The influence that predicts familiarity with nanotechnology is an individual's cultural worldview (☞ Fig. 28.12). The more individualistic a person is, the more likely he or she is to claim to know about nanotechnology, conditional on that person being simultaneously hierarchical in his outlooks – the combination of values that the experiment shows does indeed predispose individuals to become more disposed to form a positive view of nanotechnology when they learn about it.

In this study, then, we have the core of a psychometric theory of how culture influences risk perception. Individuals bear cultural predisposition toward risk – a tendency (founded on identity-protective cognition) to view some risk claims more congenial than others on the basis of latent characteristics indicated by values they share with others. This predisposition not only endows culturally diverse individuals with opposing “prior” beliefs about risk. It also decisively regulates their experience with information about the truth or falsity of those beliefs. People with opposing predispositions seek out support for their competing views through opposingly biased forms of information search. What's more, they construe or assimilate information, whatever its provenance, in opposing ways that reinforce the risk perceptions they are predisposed to form. As a result, individuals end up in a state of cultural conflict – not over values, but over *facts* – that the mere accumulation of empirical data cannot be expected readily to dispel (☞ Fig. 28.13).

## Cultural Availability

The “availability effect” describes a typical distortion that occurs when individuals assess a risk (Kahneman and Tversky 1982). If instances of some fact or contingency relevant to the risk are highly salient, individuals are more likely to notice, assign significance to, and remember them. When they are required to consider the incidence of such a contingency thereafter, the ease

with which those instances can be recalled will induce individuals to overestimate their occurrence. The ready availability of mishaps such as the Chernobyl nuclear accident, the 9/11 attack, and the Columbine school shooting massacre, for example, are thought to explain why members of the public tend to overestimate the risks of nuclear power generation, of terrorist attacks, of accidental handgun shootings, and the like, particularly in relation to less dramatic hazards – swimming pool drowning, say, or climate change.

There is a fairly obvious mystery associated with the availability effect, however: what gives one or another contingency the salience necessary to trigger the effect? The seemingly obvious answer – the vivid or horrific consequences that attend it – in fact begs multiple questions. If people viewed the accidental drowning deaths of children as being as horrific as accidental shooting deaths, then presumably they would notice (or have their attention drawn by the media to) the former more often; they would thereafter more readily recall instances of such mishaps; and as a result they would revise upward their estimation of the incidence of them relative to accidental shootings – which are in fact much rarer.

Disproportionate media coverage of various types of accidents is a common but manifestly unsatisfying explanation of the greater “availability” of them in the public mind. The media’s incentive to disproportionately cover one type of accident is itself a market-driven reflection of the public demand for news relating to that type of accident (e.g., accidental child shootings) as opposed to another, even more frequent type (accidental drownings). What is the source of that demand?

Moreover, if horrifically vivid consequences or disproportionate media coverage are what trigger availability, why do people systematically *disagree* about nuclear power plants, domestic terrorist incidents, and climate change, all of which are attended by signature images of calamity? Indeed, people disagree about the incidence of risk-relevant facts even after attending to images they *agree* are compelling and horrific: in the wake of a school shooting massacre, some people revise upward their estimate of the risks associated with permitting private citizens to own guns, while others revise upward their estimate of the risks associated with *prohibiting* law-abiding citizens from carrying guns to defend themselves. So what determines why people attach differential *significance* to salient, readily available instances of some contingency?

One possibility, I and my collaborators have conjectured, is culture (Kahan and Braman 2003a). If people are more likely to notice risk-related contingencies congenial to their cultural predispositions, to assign them significance consistent with their cultural predispositions, and recall instances of them when doing so is supportive of their cultural predispositions, then the availability effect will generate systematic individual differences among culturally diverse individuals. That would make *cultural availability* another mechanism of cultural cognition.

We examined this mechanism in a study of public perceptions of scientific expert consensus (Kahan et al. 2011). Public dispute about the extent, causes, and likely consequences of climate change often is cited as proof that substantial segments of the population are willing to buck “scientific consensus” on risk issues. But what’s the evidence that those who are skeptical of climate change believe their view is contrary to “scientific consensus?” Why not consider the possibility that such persons are conforming their impressions of what most scientists believe to their own cultural predispositions on climate change? And why not investigate whether the same is true of those who *do* perceive climate change to be a serious risk? A “cultural availability effect” would predict exactly this sort of division: If people are more likely to notice, to assign significance to, and to recall the expression of an expert opinion when it is congenial to their cultural predispositions, then they will form diametrically opposed estimates of what most

scientists believe – and not just on climate change, but on a variety of other risk issues that admit of scientific investigation but that are nonetheless culturally charged.

Our study generated two sorts of evidence suggestive of a cultural availability effect on scientific consensus. The first came from an experiment to see whether cultural predispositions affect whether someone is likely to take note of an expert's opinion. In the experiment, we asked each subject to imagine a friend was trying to make up his or her mind on the existence and effects of climate change, on the safety of nuclear power, or on the impact on crime of allowing

Geologic Isolation of Nuclear Wastes		
High Risk (not safe)	<p><b>"Using deep geologic isolation to dispose of radioactive wastes from nuclear power plants would put human health and the environment at risk.</b> The concept seems simple: contain the wastes in underground bedrock isolated from humans and the biosphere. The problem in practice is that there is no way to assure that the geologic conditions relied upon to contain the wastes won't change over time. Nor is there any way to assure the human materials used to transport wastes to the site, or to contain them inside of the isolation facilities, won't break down, releasing radioactivity into the environment.... These are the sorts of lessons one learns from the complex problems that have plagued safety engineering for the space shuttle, but here the costs of failure are simply too high.</p>	 <p>Oliver Roberts</p> <p><b>Position:</b> Professor of Nuclear Engineering, University of California, Berkeley</p> <p><b>Education:</b> Ph.D., Princeton University</p> <p><b>Memberships:</b></p> <ul style="list-style-type: none"> <li>• American Association of Physics</li> <li>• National Academy of Sciences</li> </ul>
Low Risk (safe)	<p><b>"Radioactive wastes from nuclear power plants can be disposed of without danger to the public or the environment through deep geologic isolation.</b> In this method, radioactive wastes are stored deep underground in bedrock, and isolated from the biosphere for many thousands of years. Natural bedrock isolation has safely contained the radioactive products generated by spontaneous nuclear fission reactions in Oklo, Africa, for some 2 billion years. Man made geologic isolation facilities reinforce this level of protection through the use of sealed containers made of materials known to resist corrosion and decay. This design philosophy, known as 'defense in depth,' makes long-term disposal safe, effective, and economically feasible."</p>	 <p>Oliver Roberts</p> <p><b>Position:</b> Professor of Nuclear Engineering, University of California, Berkeley</p> <p><b>Education:</b> Ph.D., Princeton University</p> <p><b>Memberships:</b></p> <ul style="list-style-type: none"> <li>• American Association of Physics</li> <li>• National Academy of Sciences</li> </ul>

Fig. 28.14

Is this an expert? In an experiment, subjects were substantially more likely to count a university professor as an "expert" when he was depicted as taking a position consistent with their own cultural predispositions on a risk issue than when he was depicted as taking a position inconsistent with the subjects' predispositions (Kahan et al. 2011)

private citizens to carry concealed handguns. The friend, we advised, was planning to buy a book to study up on the subject, but before doing so wanted the subject's advice on whether the book's author was a "knowledgeable and trustworthy expert." Subjects were shown the authors' *curriculum vitae* CV, which indicated that the author had received a Ph.D. from one elite university, was on the faculty of another, and was a member of the National Academy of Sciences. The experimental manipulation involved what the author had *written*: for each topic – climate change, nuclear power, and concealed handguns – subjects were randomly assigned a book excerpt in which the author expressed either the "high risk" or "low risk" position (Fig. 28.14). As we hypothesized, subjects were overwhelmingly more likely to find that the author was a "knowledgeable and trustworthy" expert when the author was depicted as taking a position *consistent* with the subjects' own cultural predisposition than subjects were if the author was assigned the opposing position.

If individuals are more likely to notice or to assign significance to evidence relating to expert opinion when it supports than when it contradicts their own cultural predispositions, then over time we should expect people of opposing cultural outlooks to form opposing impressions of what most experts believe. The second piece of evidence from the study showed exactly that. Polling a large representative sample of US adults, we found that culturally diverse citizens had substantially divergent perceptions of expert consensus on climate change, nuclear waste disposal, and gun control. Indeed, we found that egalitarian communitarians and hierarchical individualists also perceived scientific opinion to be different from the position taken in so-called "expert consensus reports" issued by the US National Academy of Sciences in every instance in which the NAS position differed from the one that matched the subjects' own cultural predispositions. This result seems more consistent with the conclusion that *all* segments of the population are forming culturally biased impressions of what scientists believe than that only one cares about what scientists have to say.

Under these circumstances, the availability effect will interact with cultural worldviews to generate systematic polarization on what experts believe about risk. Asked what "scientific consensus" is on climate change, on nuclear power, or on possession of concealed firearms, individuals will summon to mind all the instances they can recall of experts expressing their views and discover that the overwhelming weight of opinion favors the view consistent with her own cultural predisposition. They'll reach that conclusion, of course, only because of unconscious bias in their sampling: the fit between an expert's position and the one congenial to their cultural predisposition is what *causes* them to take note of that expert's view, to assign significance to it, and thereafter recall it. We all believe that what "most scientists think" about a risk is important. Yet we all tend to overestimate how uniformly scientists believe what we are predisposed to believe is true.

## Cultural Credibility Heuristic

Most people (in fact, all, if one thinks about it) cannot determine for themselves just how large a disputed risk is, whether of environmental catastrophe from global warming, of human illness from consumption of genetically modified foods, of accidental shootings from gun ownership, etc. They must defer to those whom they find credible to tell them which risk claims and supporting evidence to believe and which to disbelieve. The cultural credibility heuristic refers to the hypothesized tendency of individuals to impute the sorts of qualities that make an

expert credible – including knowledge, honesty, and shared interest – to the people whom they perceive as sharing their values.

One study we did that at least makes a start at testing this hypothesis focused on the HPV vaccine. HPV, or the human papillomavirus, is a sexually transmitted disease that is extremely common among young women: it's estimated that as many as 45% of those in their early 20s have been infected by it (testing permits detection only in women). It is also the leading (effectively, the *only*) cause of cervical cancer. The FDA recently approved a vaccine for females. Public health officials recommend that the vaccine be administered by age 12, before girls are likely to become sexually active, because once a female has been exposed to HPV the vaccine won't do any good. Many states in the USA are now considering legislation to require HPV vaccination as a condition of school attendance. These provisions have provoked resistances from groups who argue that vaccination, by furnishing protection against one sort of STD, will increase the incidence of unprotected sex and thus put young girls at risk of other STDs, including HIV. Opponents of mandatory vaccination also cite the risk of unanticipated, harmful side-effects from the vaccine. Numerous state legislatures have defeated legislative proposals for such programs and one state legislature, Texas's, has overridden the creation of a program created by a gubernatorial executive order.

We studied the HPV-vaccine risk perceptions of 1,500 Americans (Kahan et al. 2010b). The sample was divided into three groups. One was supplied no information about the HPV vaccine. Another was furnished balanced information in the form of opposing arguments on whether its benefits outweighed its risks.

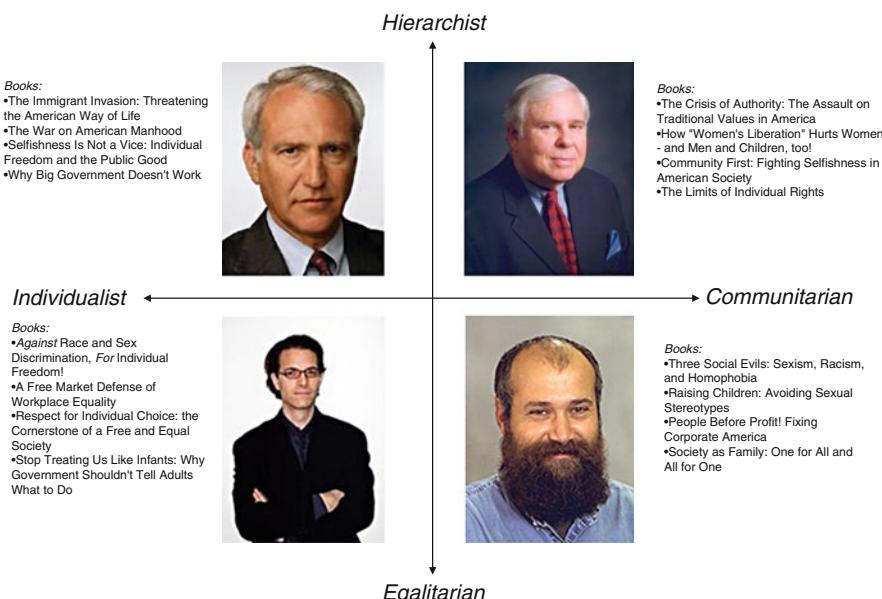
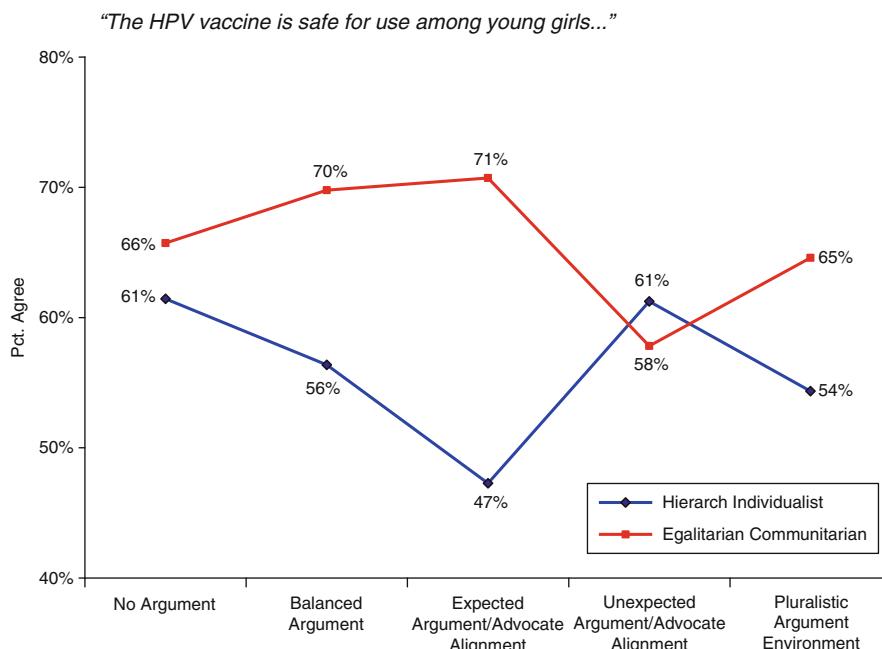


Fig. 28.15

Culturally identifiable advocates. On basis of pretests, fictional "experts" are perceived to have values characteristic of those defined by quadrants demarcated by intersection of group and grid (Kahan et al. 2010a)

The final group was exposed to the same arguments, which in this treatment were attributed to fictional, culturally identifiable experts, who were described as being on the faculties of major universities. We created the advocates in pretests. We showed one set of pretest subjects pictures of individuals, and asked them to try to guess how the pictured individuals would respond to our culture items. We then asked another group of pretest subjects to guess the fictional experts' cultural values after looking at both pictures and mock publication lists. We ended up with four culturally identifiable policy experts whose perceived cultural values located them in the quadrants defined by the intersection of group and grid (● Fig. 28.15). In the actual experiment, subjects (ones who had not previously participated in the pretests creating the culturally identifiable experts) were asked what they thought about the risks and benefits of the HPV vaccine after reading the opposing arguments, which were randomly matched with two of the four experts.

The results (● Fig. 28.16) suggested the operation of various mechanisms of cultural cognition. In the “no information” condition, there was already a division in the views of hierachal individualists and egalitarian communitarians on whether the vaccine is safe. As we had hypothesized, hierarchs and individualists were motivated to see the risks as large, the former because of their association of the vaccine with premarital sex and the latter because of



■ Fig. 28.16

**Impact of information and advocates on perceptions of HPV-vaccine risk.  $N = 1,500$ .** Gap between hierarchical individualists and egalitarian communitarians is significantly greater ( $p < 0.05$ ) in “Balanced Argument” relative to “No Argument,” and in “Expected Argument/Advocate Alignment” relative to “Balanced Argument,” it is significantly smaller ( $p < 0.05$ ) in both “Unexpected Argument/Advocate Alignment” and “Pluralistic Argument Environment” relative to “Expected Argument/Advocate Alignment”

their opposition to state-sponsored public health measures; whereas egalitarians and communitarians were motivated to see the risks as small, the former because they would see opposition as motivated by hierarchical sexual norms and the latter because they favor state-mandated public health measures. This is an identity-protective cognition effect.

These divisions grew in the “information without advocate” conditions. This is a biased assimilation and polarization effect.

In the “information with advocate” condition, the position of subjects was highly conditional on the congeniality of the experts’ values to the subjects. Where subjects received the argument they were culturally disposed to accept from an expert whose values they shared, and the argument they were culturally predisposed to reject from an expert whose values they opposed – call this the expected alignment condition – polarization grew. But where subjects received the argument they were culturally disposed to reject from the expert whose values they opposed, and the argument they were culturally predisposed to reject from the expert whose values they shared – call this the unexpected alignment condition – there was a significant decrease in polarization. Indeed, individualists and communitarians in this condition swapped places. This is powerful evidence, then, supporting the cultural credibility heuristic.

We also found out something else important. In what might be called the “pluralistic advocacy condition,” subjects observed opposing advocates whose cultural worldviews were equally proximate to or remote from their own. In this condition, polarization was also significantly diminished relative to “expected alignment” condition. In effect, confronted with a policy-advocate alignment that seemed to confound any inference that the issue was one that divided their cultural group and a competing one, individuals of diverse worldviews were less likely to polarize when they evaluated the advocates’ arguments. Presumably, this is a more realistic state of affairs to aspire to than one in which experts and arguments are aligned in a manner radically opposed to what one would expect.

Accordingly, one might identify the creation of a “pluralistic advocacy condition” – one in which risk communicators self-consciously recruit communicators of diverse cultural outlooks and are careful to *avoid* selecting ones whose identities or styles of argumentation infuse an issue with a meaning of competition or conflict between identifiable groups – as one way of *counteracting* cultural cognition. In such an environment, individuals might still disagree about the facts on risks, but they are less likely to do so along strictly cultural lines. Work on cultural cognition, then, helps to explain not only *why* we see cultural polarization on risks; it also suggests “cultural debasing” strategies – science communication techniques that make it more likely that individuals of diverse cultural outlooks will attend to information in an open-minded way.

## Cultural-Identity Affirmation

---

The next mechanism, cultural-identity affirmation, also can be seen as a type of “cultural debasing” strategy. This one is based on self-affirmation, a mechanism which is essentially the mirror image of identity-protective cognition and which has been extensively documented by Geoff Cohen, one of the Cultural Cognition Project members (Cohen et al. 2007, 2000). Identity-protective cognition posits that individuals react dismissively to information that is discordant with their values as a type of identity self-defense mechanism. With self-affirmation, individuals experience a stimulus – perhaps being told they scored high on

a test, or being required to write a short essay on their best attributes – that makes a worthy trait of theirs salient to them. This affirming experience creates a boost in a person's self-worth and self-esteem that essentially buffers the sense of threat he or she would otherwise experience while confronted with information that challenges beliefs dominant within an important reference group. As a result, individuals react in a more open-minded way to potentially identity-threatening information, and often experience a durable change in their prior beliefs.

Cultural-identity affirmation hypothesizes that you can get the same effect when you communicate information about risk in a way that affirms rather than threatens their cultural worldview. We tested this hypothesis in an experiment involving global warming (Cultural Cognition Project 2007). In the experiment, two groups of subjects all were asked to read a newspaper article that reported a study issued by a panel of scientists from major universities who found definitive evidence that the temperature of the earth is increasing, that the cause of the increase is manmade, and that the consequences of continued global warming would be catastrophic for the environment and the economy (☞ Fig. 28.17). In one treatment group, the newspaper article indicated that the study had called for the institution of stronger antipollution controls, a policy proposal *threatening* to the identity of individualists and hierarchs. In the other treatment, however, the newspaper article reported that the study had proposed removal on restrictions on nuclear power, so that American society could substitute nuclear power for greenhouse gas emitting fossil fuel energy sources. Nuclear power is *affirming* to the identity of individualists and hierarchs. Obviously, the proposed policy solution to the problem of climate change bears no logical or empirical relationship to whether the earth is heating up, whether man is causing the temperature rise, and whether global warming will have bad environmental and economic consequences. Nevertheless, consistent with the cultural-identity affirmation hypothesis, we found that individualists and hierarchs were both significantly more likely to credit the reported studying findings on these facts in the nuclear power than in the antipollution condition (☞ Fig. 28.18).

Indeed, we found that individualists and hierarchs who received the newspaper report that recommended antipollution controls were even more skeptical of the reported factual findings of the study than were individualists and hierarchs in a control group who received no newspaper story on the findings of the scientists. This is biased assimilation with a vengeance.

The practical lesson, then, is pretty clear. Don't simply bombard people with information if you are trying to make them more receptive to risks. Doing that can actually provoke a cultural-identity-protective backlash that makes certain groups even more disposed to disbelieve that the risk is a real or a serious one. Information can help, but it has to be framed in a way that *affirms* rather than threatens the cultural identities of potential risk skeptics. One way of doing that is through policy solutions that are culturally affirming of the skeptics' identities.

Or in other words, don't try to convince people to accept a solution by showing them there is a problem. Show them a solution they find culturally affirming, and then they are disposed to believe there really is a problem in need of solving.

## Further Research: Collective Management of Cultural Bias

Discussion of the last two mechanisms suggests yet another distinction between cultural cognition and various other conceptions of the cultural theory of risk. Cultural cognition suggests that the influence of worldviews on risk perceptions can be collectively managed in

## Scientific Panel Recommends Anti-Pollution Solution to Global Warming

By Jeffrey Cohen  
November 15, 2006

The American Academy of Environmental Scientists, a panel consisting of leading U.S. experts, today recommended stronger anti-pollution regulations as a response to global warming.

"Fossil fuels such as coal, natural gas, and oil are the leading cause of global warming," explained Dr. Jonathan Brasil, head of the Academy. "To reduce the volume of heat-trapping gas generated by such fuels, we recommend that the government adopt stronger anti-pollution regulations, ones strengthening ones adopted in 1970s and 1980s," Brasil said.

The group's recommendation was made in a report that examined the extent and causes of global warming and the likely consequences that would occur if global warming were not reversed.

The American Academy of Environmental Scientists, a panel consisting of leading U.S. experts, today recommended revitalization of the nation's nuclear power industry as a response to global warming.

### Highlights of AAES Report

- Fossil fuels such as coal, natural gas, and oil are the leading cause of global warming.
- Scientific evidence furnishes *irrefutable proof* of global warming. Some of the most obvious effects are visible in the Arctic, where rising temperatures and melting ice have dramatically changed the region's unique landscapes and wildlife.
- Global warming is caused by carbon dioxide and other heat-trapping gases that are emitted primarily by the burning of *fossil fuels*. These gases remain in our atmosphere for decades or even centuries.
- If it continues, global warming could have *catastrophic environmental and economic consequences*. Among the results will be extreme heat and drought, rising sea levels, and higher-intensity tropical storms. Such conditions will endanger coastal property and resources, diminish the habitability of major cities, and curtail the productivity of our farms, forests, and fisheries.
- Fossil fuels such as coal, natural gas, and oil are the leading cause of global warming. According, we strongly recommend that the government adopt *stronger anti-pollution regulations* to reduce the volume of heat-trapping gas generated by such fuels.

### Highlights of AAES Report

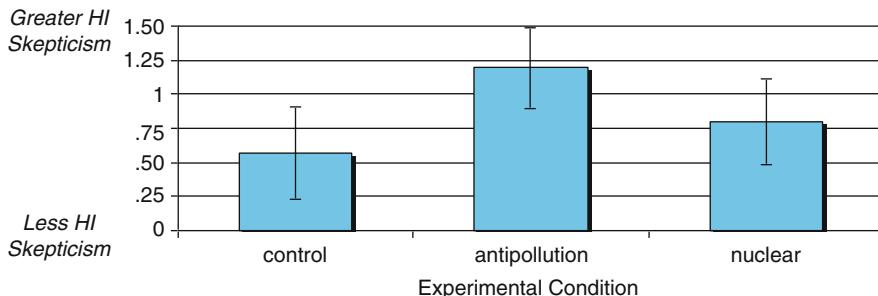
- Scientific evidence furnishes *irrefutable proof* of global warming. Some of the most obvious effects are visible in the Arctic, where rising temperatures and melting ice have dramatically changed the region's unique landscapes and wildlife.
- Global warming is caused by carbon dioxide and other heat-trapping gases that are emitted primarily by the burning of *fossil fuels*. These gases remain in our atmosphere for decades or even centuries.
- If it continues, global warming could have *catastrophic environmental and economic consequences*. Among the results will be extreme heat and drought, rising sea levels, and higher-intensity tropical storms. Such conditions will endanger coastal property and resources, diminish the habitability of major cities, and curtail the productivity of our farms, forests, and fisheries.
- Fossil fuels such as coal, natural gas, and oil are the leading cause of global warming. According, we strongly recommend that the government adopt *stronger anti-pollution regulations* to reduce the volume of heat-trapping gas generated by such fuels.
- If it continues, global warming could have *catastrophic environmental and economic consequences*. Among the results will be extreme heat and drought, rising sea levels, and higher-intensity tropical storms. Such conditions will endanger coastal property and resources, diminish the habitability of major cities, and curtail the productivity of our farms, forests, and fisheries.
- Fossil fuels such as coal, natural gas, and oil are the leading cause of global warming. According, we strongly recommend that the government adopt *stronger anti-pollution regulations* to reduce the volume of heat-trapping gas generated by such fuels.

The group's recommendation was made in a report that examined the extent and causes of global warming and the likely consequences that would occur if global warming were not reversed.

The group's recommendation was made in a report that examined the extent and causes of global warming and the likely consequences that would occur if global warming were not reversed.

Fig. 28.17

Alternative versions of newspaper article reporting scientific findings on climate change. Subjects were assigned to one of three conditions, in two, they read one of the featured newspaper articles, each of which summarized scientific findings, in the red inset, indicating that climate change was occurring, was caused by humans, and was likely to have catastrophic effects unless reversed. Subjects in the third, control condition read a newspaper article about debate on the need for a new traffic signal at a busy intersection (Cultural Cognition Project 2007)

**Fig. 28.18**

**Effect of framings on hierarchical-individualist skepticism on global warming facts.**  $N = 500$ . Bars denote the size of the difference between egalitarian and hierarchical individualist subjects' beliefs in climate change facts, as measured by z-scores on a composite scale that combined responses to questions on whether climate change is occurring, whether it is caused by human activity, and whether it would have adverse environmental impacts if not contained or reversed. Derived from linear multivariate regression in which worldviews and experimental condition were treated as predictors. Confidence intervals reflect 0.95 level of confidence (Cultural Cognition Project 2007)

a manner that simultaneously advances the interests of persons of all cultural persuasions. Identifying the means by which this end can be realized, I'm convinced, should be regarded as the priority of future cultural-cognition research.

The meta-worldview of Douglas and Wildavsky features necessary and permanent cultural conflict. Because there is no culture-free perspective, it is not possible for individuals to "overcome" reliance on their worldviews in apprehending risks. As a result, it is not possible for society to overcome the persistent struggle of opposing cultural groups to designate forms of behaviors associated with their rivals as sources of danger that must be repressed. One or another group might gain the upper hand, and thus impose its view, at least for a time. One might even form a temporary rooting interest for one or another on grounds that are seemingly utilitarian in nature (Wildavsky 1991). But the idea of brokering a peace between them – of formulating a positive-sum outcome to their bitter competition – would seem to defy the logic of cultural theory.

Cultural cognition is more catholic. Nothing in its account of the mechanisms that connect culture to risk perceptions implies that those dynamics are exclusive of others that might inform individual apprehension of risk. Nor does anything in that account entail that the contribution that alternative cultural worldviews make to risk perception are static and relentlessly oppositional.

This stance, then, creates the possibility, at least as a matter of theory, that adherents to competing ways of life might converge on shared understandings of societal risk and the most effective means for abating them. One strategy for promoting such an outcome involves the adroit framing of information, and of policies, to make them bear a plurality of meanings that can be simultaneously endorsed by opposing cultural groups. There are seeming historical examples of this dynamic – ones involving convergence of cultural groups on environmental policies in the USA and abortion policy in France, for example – which my collaborators and I call "expressive" or "social meaning overdetermination" (Kahan and Braman 2006; Kahan 2007).

Working within the logic of cultural cognition, we have tried to systematize “social meaning overdetermination” as a strategy for generating positive-sum solutions to cultural conflicts in political life. At least one other group of scholars working within the broad outlines of “group–grid” theory have proposed a similar approach, which they call “clumsy solutions.” (Verweij and Thompson 2006).

Another strategy, one unique to cultural cognition and reflecting its emphasis on mechanisms, suggests the value of structuring democratic deliberation in ways that effectively lessen participants’ *reliance* on culture. Many of the mechanisms of cultural cognition involve the use of cultural cues as a heuristic or mental shortcut. But as experimental studies show, it’s possible to disable or blunt culture’s heuristic influence: when people’s cultural identities are affirmed, they don’t experience the threatening affective response, or are less influenced by it, as they consider information that challenges beliefs that predominate in their group; when they can’t discern a consistent connection between the cultural identity of advocates and positions on some risk issue, they can’t simply adopt the position of the advocate whom they perceive as having values most like theirs. At least in theory, then, it should be possible to build into policymaking institutions and procedures devices that similarly stifle the sorts of cues that the mechanisms of cultural cognition depend on. When that happens, individuals will be forced to process information in a different way, maybe in a more considered way, or maybe in a way that reflects other cues that are reliable but not culturally valenced. In the resulting deliberative environment, individuals might not immediately converge on one set of factual beliefs about risks and risk mitigation. But they won’t spontaneously split into opposing cultural factions on those matters.

If there is a meta-worldview for cultural cognition, it is that this state of cultural *depolarization* is a good thing (Kahan 2007). It’s a good thing, to begin, morally speaking. Because culturally infused disagreements over global warming, gun control, vaccination of school girls for HPV, and the like *is* experienced by all as a form of conflict between contesting cultural factions, the polarizing effect of cultural cognition poses a distinctive sort of threat to liberal political life.

But neutralizing cultural polarization is also a good thing instrumentally speaking. Nothing in cultural theory, as Wildavsky and Douglas originated it or as it has been refined thereafter, implies that there are no real facts about risk or that we can’t, through the best information we can discover on the workings of our world, form better or worse understandings of those facts. Regardless of their differences about the ideal society, hierarchs and egalitarians, communitarians and individualists all have a stake in policymaking being responsive to that information; regardless of what lives they want to live, they can all live those lives better when their health is not threatened by contamination of their food or air, when their economy is insulated from disruptive influences (including governmental ones), and when they are free of domestic and external security threats. These culturally diverse citizens then would presumably all agree that they ought to structure their political institutions and processes in a manner that counter cultural polarization on issues of risk, because when society is culturally polarized the best understandings we have about risk are *less* likely to become operative as *soon* as they would otherwise.

Or at least they sometimes would agree, behind a cultural veil of ignorance as it were, that that’s what they want. The possibility of self-consciously managing cultural cognition presents a host of complicated moral questions in large part because the moral status of beliefs we form as a result of our cultural identities is complicated. Only a fool – a moral idiot – would regard those beliefs as *uniformly* unworthy of his or her endorsement. We don’t *just* want to live or

even live well (in a material sense); we also want to live virtuously and honorably. There's no way to figure out how to do that, I am convinced, without certain moral insights that *only* cultural infused modes of perception can afford (Kahan 2010).

So part and parcel of any project to manage cultural cognition is an informed moral understanding of when, as individuals and as a democratic society, we *should* be responsive to cultural cognition and when we *shouldn't*. Like Douglas and Wildavsky, I don't think it makes any sense to believe we can "overcome" our cultural commitments in considering that issue. But I see nothing that makes me believe that persons of diverse cultural persuasions will inevitably be driven into irreconcilable disagreement over how to resolve it. Until I *am* presented with evidence that forces me to accept that sad conclusion, I will pursue a conception of cultural theory that sees dissipating conflict over risk as the very point of explaining it.

## Acknowledgments

---

Research described herein was supported by the National Science Foundation (Grants SES-0621840 & SES-0242106) and the Oscar M. Ruebhausen Fund at Yale Law School.

## References

---

- Aiken LS, West SG, Reno RR (1991) Multiple regression: testing and interpreting interactions. Sage, Newbury Park
- Balkin JM (1998) Cultural Software. Yale Univ. Press, New Haven
- Boudon R (1998) Social mechanisms without black boxes. In: Hedström P, Swedberg R (eds) Social mechanisms: an analytical approach to social theory. Cambridge University Press, Cambridge, pp 172–203
- Braman D, Kahan DM, Grimmelmann J (2005) Modeling facts, culture, and cognition in the gun debate. Soc J Res 18:283–304
- Cohen GL (2003) Party over policy: the dominating impact of group influence on political beliefs. J Pers Soc Psychol 85(5):808–822
- Cohen GL, Aronson J, Steele CM (2000) When beliefs yield to evidence: reducing biased evaluation by affirming the self. Pers Soc Psychol Bull 26(9): 1151–1164
- Cohen GL, Bastardi A, Sherman DK, Hsu L, McGoey M, Ross L (2007) Bridging the partisan divide: self-affirmation reduces ideological closed-mindedness and inflexibility in negotiation. J Pers Soc Psychol 93(3):415–430
- Cultural Cognition Project (2007) The second national risk and culture study: Making sense of - and making progress in - the American culture war of fact. Yale Law School, New Haven, Conn. October 2007, available at <http://www.culturalcognition.net/projects/second-national-risk-culture-study.html>
- Dake K (1990) Technology on trial: orienting dispositions toward environmental and health standards. Ph.D. dissertation. University of California at Berkeley, Berkeley
- Dake K (1991) Orienting dispositions in the perception of risk: an analysis of contemporary worldviews and cultural biases. J Cross Cult Psychol 22:61
- Dake K (1992) Myths of nature: culture and the social construction of risk. J Soc Issues 48(4):21–37
- Deuteronomy 28:22 (1997). In New American Standard Bible
- Douglas M (1966) Purity and danger: an analysis of concepts of pollution and taboo. Routledge, London
- Douglas M (1970) Natural symbols: explorations in cosmology. Barrie & Rockliff the Cresset Press, London
- Douglas M (1982) In the active voice. Routledge & K. Paul, London/Boston
- Douglas M (1985) Risk acceptability according to the social sciences. Russell Sage, New York
- Douglas M (1986) How Institutions Think, 1st edn. Syracuse University Press, Syracuse
- Douglas M (1992) Risk and blame: essays in cultural theory. Routledge, London/New York
- Douglas M (1997) The depoliticization of risk. In: Ellis RJ, Thompson M (eds) Culture matters: essays in

- honor of Aaron Wildavsky. Westview Press, Boulder, pp 121–132
- Douglas M (2003) Being fair to hierarchists. *Univ Penn Law Rev* 151(4):1349–1370
- Douglas M, Wildavsky AB (1982) Risk and culture: an essay on the selection of technical and environmental dangers. University of California Press, Berkeley
- Ellis RJ, Thompson F (1997) Culture and the environment in the Pacific Northwest. *Am Polit Sci Rev* 91(4):885–898
- Elster J (1985) Making sense of Marx. Cambridge University Press, Cambridge
- Finucane M, Slovic P, Mertz CK, Flynn J, Satterfield TA (2000) Gender, race, and perceived risk: the “white male” effect. *Health Risk Soc'y* 3(2):159–172
- Gastil J, Jenkins-Smith H, Silva C (1995) Analysis of cultural bias survey items (Institute for Public Policy, University of New Mexico, 1995)
- Giner-Sorolla R, Chaiken S (1997) Selective use of heuristic and systematic processing under defense motivation. *Pers Soc Psychol Bull* 23(1):84–97
- Gross JL, Rayner S (1985) Measuring culture: a paradigm for the analysis of social organization. Columbia University Press, New York
- Jenkins-Smith H (2001) Modeling stigma: an empirical analysis of nuclear waste images of Nevada. In: Flynn J, Slovic P, Kunreuther H (eds) Risk, media, and stigma: understanding public challenges to modern science and technology. Earthscan, London/Sterling, pp 107–132
- Jenkins-Smith HC, Herron KG (2009) Rock and a hard place: public willingness to trade civil rights and liberties for greater security. *Polit Policy* 37(5): 1095–1129
- Judd CM (2000) Everyday data analysis in social psychology: comparisons of linear models. In: Reis HT, Judd CM (eds) Handbook of research methods in social and personality psychology. Cambridge University Press, New York, pp 370–392
- Kahan DM (2007) The cognitively illiberal state. *Stan L Rev* 60:115–154
- Kahan DM (2010) Emotion in risk regulation: competing theories. In: Roeser S (ed) Emotions and risky technologies. Springer, Dordrecht
- Kahan DM, Braman D (2003a) Caught in the crossfire: a defense of the cultural theory of gun-risk perceptions - response. *Univ Penn Law Rev* 151(4): 1395–1416
- Kahan DM, Braman D (2003b) More statistics, less persuasion: a cultural theory of gun-risk perceptions. *Univ Penn Law Rev* 151:1291–1327
- Kahan DM, Braman D (2006) Cultural cognition of public policy. *Yale J L & Pub Pol'y* 24:147–170
- Kahan DM, Slovic P, Braman D, Gastil J (2006) Fear of democracy: a cultural critique of sunstein on risk. *Harv Law Rev* 119:1071–1109
- Kahan DM, Braman D, Gastil J, Slovic P, Mertz CK (2007) Culture and identity-protective cognition: explaining the white-male effect in risk perception. *J Empirical Legal Stud* 4(3):465–505
- Kahan DM, Braman D, Slovic P, Gastil J, Cohen G (2009) Cultural cognition of the risks and benefits of Nanotechnology. *Nat Nanotechnol* 4(2):87–91
- Kahan DM, Braman D, Cohen GL, Gastil J, Slovic P (2010a) Who fears the HPV vaccine, who doesn't, and why? An experimental study of the mechanisms of cultural cognition. *Law Hum Behav* 34:501–16
- Kahan DM, Braman D, Monahan J, Callahan L, Peters E (2010b) Cultural cognition and public policy: the case of outpatient commitment laws. *Law Hum Behav* 34:118–140
- Kahan DM, Jenkins-Smith HC, Braman D (2011) Cultural cognition of scientific consensus. *J Risk Res* 14:147–74
- Kahneman D, Tversky A (1982) Availability: a heuristic for judging frequency and probability. In: Kahneman D, Slovic P, Tversky A (eds) Judgment under uncertainty: heuristics and biases. Cambridge University Press, Cambridge/New York, pp 163–178
- Langford IH, Georgiou S, Bateman IJ, Day RJ, Turner RK (2000) Public perceptions of health risks from polluted coastal bathing waters: a mixed methodological analysis using cultural theory. *Risk Anal* 20(5):691–704
- Lord CG, Ross L, Lepper MR (1979) Biased assimilation and attitude polarization – effects of prior theories on subsequently considered evidence. *J Pers Soc Psychol* 37(11):2098–2109
- Mamadouh V (1999) Grid-group cultural theory: an introduction. *GeoJournal* 47:395–409
- Manton KG, Woodbury MA, Stallard E, Corder LS (1992) The use of grade-of-membership techniques to estimate regression relationships. *Sociol Methodol* 22:321–381
- Marris C, Langford IH, O'Riordan T (1998) A quantitative test of the cultural theory of risk perceptions: comparison with the psychometric paradigm. *Risk Anal* 18(5):635–647
- Nisbett RE (2003) The geography of thought: how Asians and Westerners think differently—and why. Free Press, New York
- O'Connor RE, Bord RJ, Fisher A (1998) Rating threat mitigators: faith in experts, governments, and individuals themselves to create a safer world. *Risk Anal* 18(5):547–556
- Peters E, Slovic P (1996) The role of affect and worldviews as orienting dispositions in the perception and

- acceptance of nuclear power. *J Appl Soc Psychol* 26(16):1427–1453
- Rayner S (1992) Cultural theory and risk analysis. In: Krimsky S, Golding D (eds) *Social theories of risk*. Praeger, Westport, pp 83–115
- Silva CL, Jenkins-Smith HC (2007) The precautionary principle in context: U.S. and E.U. scientists' prescriptions for policy in the face of uncertainty. *Soc Sci Q* 88(3):640–664
- Slovic, P (2000) *The Perception of Risk*. Earthscan Publications, London/Sterling, VA
- Sjöberg L (1998a) Explaining risk perception: an empirical evaluation of cultural theory. In: Löfstedt RE, Frewer L (eds) *The earthscan reader in risk and modern society*, vol 2. Earthscan, London, pp 115–132
- Sjöberg L (1998b) World views, political attitudes, and risk perception. *Risk Health Saf Environ* 9:137–152
- Thompson M, Ellis R, Wildavsky A (1990) *Cultural theory*. Westview Press, Boulder
- Verweij M, Thompson M (eds) (2006) *Clumsy solutions for a complex world: governance, politics, and plural perceptions*. Palgrave Macmillan, Hounds-mills/Basingstoke/Hampshire/New York
- Verweij M, Douglas M, Ellis R, Engel C, Hendriks F, Lohmann S et al (2006) Clumsy solutions for a complex world: the case of climate change. *Public Admin* 84(4):817–843
- Wildavsky AB (1991) *The rise of radical egalitarianism*. American University Press, Washington, DC
- Wildavsky A, Dake K (1990) Theories of risk perception: who fears what and why? *Daedalus* 114:41–60



# 29 Tools for Risk Communication

*Britt-Marie Drottz-Sjöberg*

Norwegian University of Science and Technology (NTNU), Trondheim,  
Norway

<b><i>Introduction</i></b> .....	<b>762</b>
<b><i>Background</i></b> .....	<b>765</b>
<b><i>Current Research</i></b> .....	<b>771</b>
Social Challenges: An Example of Public Reactions to Inviting a Geological Investigation .....	771
Group Challenges: An Example of Concerns of Politicians and Nuclear Experts .....	772
Challenges of Contents of Risk Information: An Example of Localization of a LNG Storage Tank .....	774
Challenges of Contents of Risk Information: An Example of Safety Data Sheets .....	776
International Challenges: Examples from the ARGONA Project .....	777
<b><i>Further Research</i></b> .....	<b>781</b>

**Abstract:** This chapter exemplifies risk communication projects and aims at illustrating the specificity of each risk communication task as well as what general conclusions can be gained from such work. Tools for risk communication include the selection of strategic and theoretical approaches as well as a number of considerations in applied settings. The examples show the influence of social and historic events that embed a communication setting or a conflict situation resulting in a risk communication project, and how values, attitudes, and feelings steer thinking and behavior in intergroup interactions. The examples also generate some ideas for the improvement of risk communication work. The RISCOM model of transparency is presented shortly, although most described projects were based on the less elaborate designs of group discussions. It is concluded that risk communication always takes place in a social setting, involving various interests, power relations, and actors' own agendas, but that the aim to communicate about specific risks nevertheless can be focused on clarification, understanding, and learning. The described tools for risk communication aim at achieving clarity in dialogues characterized by openness and interaction regarding risk issues to enhance problem solving and democracy.

## Introduction

---

Communication is vital for all aspects of life. In addition to every individual's personal communication experiences, there are many academic fields that systematically describe the nature of communication and communicative processes. A review of such theories is not the approach of this chapter, but examples can be found in several excellent contributions in this handbook. From the point of view of describing "tools for risk communication," the early and well-known question by Lasswell (1948) – "Who says what to whom with what effect?" – may be a good enough starting point. Risk communication projects are often initiated when some risk aspect has already been highlighted, often as a controversial issue, and when problems have surfaced in the interaction between parties that hold different views, or when there is a strong suspicion that serious conflict is approaching due to localization of an establishment, e.g., in the "Not in my backyard" (NIMBY) attitude context (see, e.g., Dunlap et al. 1993).

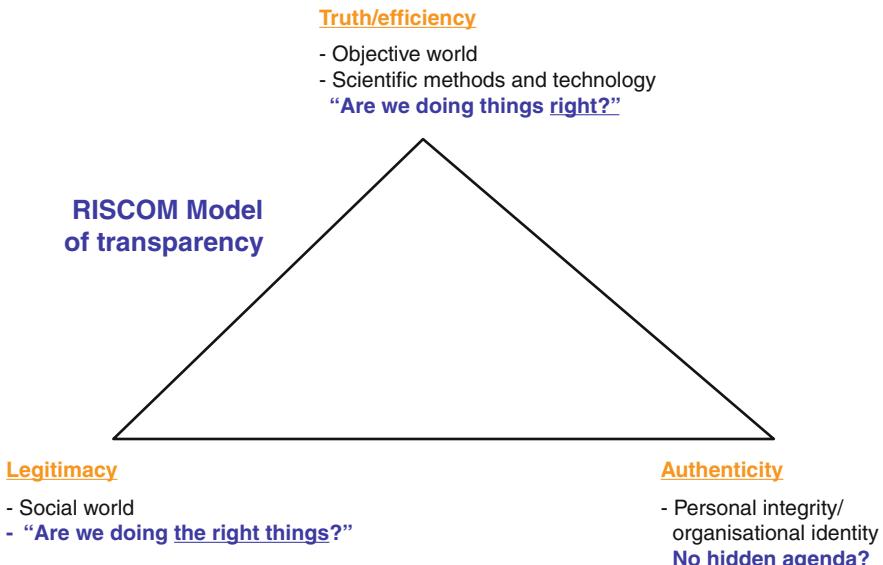
There are obvious challenges in the work, but also not so obvious challenges, and the latter may constitute the real challenges in practical work. Research literature on communication usually provides composite theoretical frameworks, and interaction models including senders and receivers, turn taking in dialogues, and elaborations on the cognitive processes of interpretation and understanding (e.g., Krauss and Fussell 1996; Littlejohn 1999). The necessity of presenting clear messages using common words is also often underlined in risk communication. Thus, these ordinary communication aspects are not unimportant. However, even if a communication situation meets obvious interaction requirements it may present challenges of different kinds. One of those is the situation where one or several persons in an exchange actually do not understand that he or she does not understand. It may take some time before highly specialized experts, or excited protagonists, come around to ask, e.g., "What do you mean with that?" or "How do you define that?" Misunderstandings can easily result from different conceptualizations of terms within specialization areas, lack of knowledge, or simply be based on unwillingness to consider an opponent's point of view. If a misunderstanding is not observed at all, the involved will keep their own interpretations of the exchange and may later discover that others act in ways that are inconsistent or contradictory to how they had understood the

results of the discussion. If such a development takes place and is not rectified, it is not uncommon that those involved lose interest and trust in each other. If a misunderstanding is not observed, or not corrected, the interacting parties might choose to disregard one another's comments, leave the communication situation in frustration, or try to gather support from others for the own perspective, and thus escalate a misunderstanding into a social conflict.

A related unobtrusive challenge is the type of situation that involves the determining of whether or not a misinterpretation was intentional or not. Words and concepts are often laden with multiple meanings, and therefore involve ambiguities, and it is not uncommon to misinterpret sincerely produced statements. Skilled demagogues or change agents, however, are talented in the choice of words and ambiguities so that they favor their own goals or rhetorical points. Since risk communication projects often are developed on the basis of already existing group polarizations or social conflicts, and are more often introduced reactively than proactively, it is of extreme importance to approach such work with a clear understanding of not only the intricate communication challenges but also of the simultaneous social interaction processes that frame the situation that involve individuals and groups.

Risk communication work is much trickier than one would expect at first glance. The definition of what is a hazard, or a risk, is not unproblematic and ordinarily a communication process also involves many social, interpersonal, and personality aspects. In addition to *defining* the relevant risk issues in a certain *social situation*, risk communication has the main goal to *capture* and *discuss* the particular risk issues around which a project is based. "Risk" will here be defined in terms of perceived risk, i.e., a phenomenon evaluated by an individual to have some kind of negative effect or consequence attached to it. However, perceptions and subjective evaluations can, for many reasons, be erroneous. For example, they can be based on selective or incorrect information, inadequate knowledge, and be more or less influenced by feelings and emotions. Therefore, communicating about risk is related to investigating and clarifying the correct or the optimal answer based on perceived risk and available information. If enough valid knowledge about a risk can be made available, the risk communication task can be focused on elucidating that knowledge in a dialogue process aiming at increased understanding and management of the risk. Risks are communicated by words as well as by numbers and symbols. Thus, the risk communication task is related to clarifying concepts and estimates from risk or safety analyses, i.e., quantitative calculations of the probabilities of the involved negative outcomes. For example, a local residence area has received the information that the communal water supply contains a health risk, i.e., a certain type of parasite. The perceptions of this particular risk will differ, due to, e.g., the information given, previous relevant knowledge and experience, responsibility for others, personal health status, etc. Risk communication, in this water contamination example, will involve an exchange of various perceptions, and different types of reactions, but also expert estimates of the health risk to various categories of water users or groups of individuals, and regarding different uses of the water such as drinking or washing. Thus, risk communication goes beyond ordinary exchanges of personal views on a subject matter since it also aims to communicate to what degree there is a danger, to whom, and what can be done to avoid or to mitigate negative consequences.

This chapter gives examples of the layers of complexity in communicating about risk. As noted above, risk communication projects are often started when there is already an explicit problem, often highlighted by very different interpretations or perceptions of a risk or situation. Such situations are often characterized by severed trust between the interacting parties. It may therefore be the case that sources that ordinarily are judged as providing valid



**Fig. 29.1**  
The RISCOM model of transparency

information, e.g., expertise, authorities, or decision-makers in various contexts, no longer are trusted, putting an information recipient in a situation of uncertainty, discomfort, and maybe worry. Without sufficient trust in information sources, a person may not be willing to apply the everyday “rules of thumb” that facilitate choices and behavior. Instead, much more deliberate thinking is required to sort out premises, options, and actions. This step is difficult because it involves the application of knowledge gathering skills, time to process new information, and motivation. It is in this phase of active information processing that risk communication projects may be helpful. However, people react differently to challenges and dangers. Some try to find answers, others seek comfort among similarly inclined, and yet others may instead prefer not to be reminded of the hazard or simply deny a risk. One challenge in a risk communication project is therefore to provide a neutral arena that allows the voicing of different perceptions, opinions, and reactions but which, at the same time, strives at clarification of the risk issue.

A theoretical framework that incorporates the simultaneous communication of scientific facts or expertise information, social norms, and personal characteristics is the risk communication model of transparency, RISCOM, developed and used by Andersson and colleagues (Andersson et al. 1998, 1999, 2006). It is based on Jürgen Habermas’ writings on communicative action (e.g., Habermas 1988) that claim that in order for an action to be communicative the statements must be true, right, and truthful. That is to say that every competent speaker raises three claims: about the “objective world,” i.e., “What I say is true” (theoretical interest, scientific quest), about the “social world”, i.e., “What I say is legitimate” (practical interest, societal norms), and about an “inner world,” i.e., “I am truthful and speak from conviction” (emancipatory interest, values held by individuals or groups). The RISCOM triangle is presented in Fig. 29.1.

The RISCOM model was developed in a number of projects in the 1990s, and it has been tested in European Union projects in the 2000s (ARGONA, IPPA). The RISCOM model of

transparency is based on four principles: a multi-perspective starting point, “stretching capacity,” impartiality and fairness, and the public sphere as the working arena, i.e., the *transparency forum*. The concept of transparency is defined as an outcome of an ongoing learning process in a given policy area, which increases the participants’ or stakeholders’ appreciation of related issues and provides them with channels to challenge, or to *stretch*, other participants’ positions regarding, e.g., to meet requirements for technical explanations, proof of authenticity, and legitimacy of actions. Transparency, in this context, requires a *guardian* to secure process integrity. The transparency forum includes a *reference group*.

The reference group builds on stakeholder participation, e.g., industry, academia, authorities, NGOs, and is established by a formal agreement between the participants. This reference group agrees on a structure for the communication process, i.e., defines the “levels of meaningful dialogue.” It also arranges the overall process, including seminars, hearings, etc. The second phase involves the communication activities, e.g., a hearing, seminars, and group work, in a tailored format made as public as possible and preferably with media attendance. The third phase involves documentation and dissemination of results. The transparency forum does not end with recommendations to decision-makers. The sole aim is to create an arena where all stakeholders increase their awareness and learning. After that, the ordinary political system takes over. In the context of this chapter and the examples given below, i.e. the ARGONA project was designed in line with these ideas.

The approach to *tools* for risk communication in this chapter, attempts to provide a basis for enhanced understanding of the challenges in communication about risk and what could be useful approaches, or tools, in those situations that need to consider both context and content. Risk communication projects are always situated in a larger social and political context, and core beliefs and values acting in that larger context will seep into individual views and group discussions. However, a risk communication project also has a specified task. That task defines the central content. The context and the content interact in many ways. This chapter presents examples of challenges and lessons learnt from a few selected risk communication projects.

## Background

---

Communication has, on the one hand, been described as an interaction impossible not to perform, i.e., “one cannot not communicate” (Watzlawick et al. 1967), and on the other hand as mere transmission of information which is not necessarily received or understood (for theories of human information processing, persuasion, and communication see, e.g., Petty and Cacioppo 1986; Petty and Wegener 1998; Littlejohn 1999; McQuail 2000; Chaiken 2003). Very often, we believe that a sent message was clear and unambiguous although responses and reactions may later indicate totally different and unexpected interpretations. Such unintended communication exchange is quite common in ordinary life. In person-to-person exchanges, it is usually possible to correct mistakes during the interaction, but, of course, that requires mutual interest and willingness to listen. Imagine a situation involving reluctance to listen in combination with the intricate challenges in communicating specific definitions, or the intended meaning, of abstract concepts such as “hazard,” “risk,” and “uncertainty.” Furthermore, imagine discussing these basic concepts in situations where people feel unjustly exposed to risk and fear for their health or safety or in a situation where various groups actively pursue campaigns to achieve specific goals, e.g., exaggerating or minimizing threats in order to

influence a decision. The examples highlight the importance of pedagogical skills as well as the need to achieve mutual respect to enable dialogue and understanding in risk communication projects.

There are excellent reviews of the history, contents, and development of the research field elsewhere (Plough and Krimsky 1987; Covello et al. 1989; Fischhoff 1995; Renn 1998; North 1998; Fischhoff et al. 1993; Boholm 1998; Gurabardhi et al. 2005; Wright et al. 2006; Breakwell 2007) and here it is suffice to say that the risk communication area has very rapidly gained in scope and content, especially since the 1990s. Today it represents a well-established field of research with contributed knowledge and experience from many scientific disciplines, civic organizations, and other contributors. It presents numerous theoretical models and risk communication applications related to highlighted societal challenges and events, e.g., faulty consumer products, chemical contamination, radioactive fallout, oil spills, nuclear waste, GM crops, utilization of gene and nanotechnologies, etc. (e.g., Dunlap et al. 1993; Levidow et al. 2000; Kennedy et al. 2010). There are also numerous studies related to local concern and conflict attached to existence, development, or change of infrastructure and industrial production (e.g., Moffatt et al. 2003; Prades et al. 2009; Kari et al. 2010).

It is not possible to present a recipe for success of a risk communication project. However, there are many risk communication failures to learn from which, together with common sense, can help steer strategies and practical considerations between the rock and the hard place. Thus, it is easier to predict failure than it is to predict success and some indicators of failure or initiatives leading to negative results are exemplified below. Large projects affect a number of people and may involve many actors or interest groups. Comments from participants usually include reflections on the design of the process as much as its content. It should also be mentioned, however, that change in itself is perceived to involve more uncertainty than is an “undisturbed” status quo. Suggestions involving change may therefore be perceived as more risky than suggestions aiming at refraining from action. Thus, a risk communication project will have to handle reactions to change as well as to the process design and the specific risk issue content.

The road to failure involves not giving any, or sufficient, information to the directly affected or the general public before an important event or decision takes place. Lack of early notification before an important event, forgetting to mention, or to strategically withhold information regarding plans or decisions related to an important issue, cause discontentment. An example could be to execute plans of investigating a municipality for uranium ore without prior notice and open debate, or to start an expansion of an existing industrial establishment without prior information. Another golden road to problems is to allow, or participate in, non-transparent decision processes, or to be obscure or ambiguous regarding responsibilities so that information about what, when, how, and by whom is not clear. To increase the level of irritation one can classify, or treat as confidential central documents, such as detail plans or results of risk analyses. Other options involve hindering insight in an ongoing decision-making process by blocking, or otherwise not making available, communication channels for information, questions, and responses. It is also usually helpful in achieving public outrage or distrust to force a decision, or complete a political decision process, before participation processes or a public debate have been initiated or sufficiently thorough.

With respect to risk events, or projects in local communities that specifically concern local groups of inhabitants, one path to irritation and possible future failure goes via media, i.e., to allow the first information about a change or decision to reach the directly concerned groups through ordinary news and media channels. It is a failure path to avoid early contacts with

central stakeholders, and to insist, e.g., via media, that there is no risk at all, or at least none that is known so far. If this is not enough to stir inflamed debate, an actor can certainly also make sure there is no available service arrangement responding to public inquiries, by phone, letter, or electronic media.

If a communication or participation process is underway, a failure predictor includes to avoid or disregard the most creative or farfetched “worst-case” scenarios that critics come up with. A response saying that such events will almost certainly not happen so why waste time and energy in looking into the matter, will quickly impact the process. Also, to avoid responding to inquiries about rumors or examples of other actual or possible mishaps or accidents usually increase the interaction temperature. Generally, to approach individuals and groups as if they were totally ignorant and without any possibility to understand the extreme complexity of the issue at hand usually helps create major communication problems. However, avoiding these paths leading to communication failure and maybe social disruptions does not necessarily guarantee success. Successful approaches instead seem to involve active involvement, early communicative approaches, comprehensive strategies, and systematic work to achieve success instead of mere clinging to strategies of avoiding problems.

It must be added here, of course, that communicative processes described in this chapter are embedded in social systems based on democratic values. If a society does not adhere to values of equality, free speech, and political participation, etc., or have not accepted international conventions or, e.g., EU directives on safety and information (cf. the [Seveso II directive](#); see also De Marchi 1991; De Marchi and Funtowicz 1994; Stern and Fineberg 1996), then the content of this chapter is not applicable. Thus, the cases presented further below aim at illustrating some challenges that may be encountered in risk communication processes related to vital societal projects conducted in democratic societies.

Access to information and the right to voice one’s opinion are basic principles in democratic societies. Rather recent developments, of special relevance to risk communication projects, include the opening up for concerned persons to *participate* in decision processes. For example, the Aarhus convention (1998) with specific relevance to environmental issues builds on earlier declarations related to humans, environment, and development. It states, e.g., that the parties to the convention (Convention on Access to Information, Public Participation in Decision-Making and Access to Justice in Environmental Matters, Aarhus, Denmark, June 25, 1998. See <http://www.unece.org/env/pp/documents/cep43e.pdf>) are

- ▶ *Recognizing that, in the field of the environment, improved access to information and public participation in decision-making enhance the quality and the implementation of decisions, contribute to public awareness of environmental issues, give the public the opportunity to express its concerns and enable public authorities to take due account of such concerns*

*Aiming thereby to further the accountability of and transparency in decision-making and to strengthen public support for decisions on the environment, . . .*

The following quote from later work underlines the importance of public participation processes (“Vision and Mission” of the Aarhus Convention Strategic Plan, paragraph 4, adopted by the Meeting of the Parties to the Aarhus Convention, in Riga, Latvia, on 13 June, 2008. <http://www.unece.org/env/pp/>): “The serious environmental, social and economic challenges faced by societies worldwide cannot be addressed by public authorities alone without the involvement and support of a wide range of stakeholders, including individual citizens and civil society organizations.”

Environmental issues directly affecting health and safety, and issues that involve the continuous interactions of humans, the environment, and societal and technological developments, make the Aarhus convention and similar international declarations or conventions, e.g., Agenda 21 from the 1992 Rio conference, basic guides to values enhanced in risk communication work. In 1985, the European Union introduced the *Environmental Impact Assessment Directive* ([EIA, 85/337/EEC](http://eia.85/337/EEC)) for analyses of environmental effects of certain planned activities. A summary provided on the Internet states that the EIA “must identify the direct and indirect effects of a project on the following factors: man, the fauna, the flora, the soil, water, air, the climate, the landscape, the material assets and cultural heritage, and the interaction between these various elements.” ([http://europa.eu/legislation\\_summaries/environment/general\\_provisions/128163\\_en.htm](http://europa.eu/legislation_summaries/environment/general_provisions/128163_en.htm)) The EIA includes consultation processes with locally affected and other stakeholders. A 25-year anniversary EU conference on the theme “Successes – Failures – Prospects” was held in 2010 for evaluation and discussion of the directive.

The European *Strategic Environmental Assessment*, the SEA Directive 2001/42/E, also known as strategic environmental impact assessment, applies to an authority (at national, regional, or local level) and relates to a wide range of public programs and plans. It requires, e.g., consultation with environmental authorities and assessment of reasonable alternatives. The European Commission (<http://ec.europa.eu/environment/eia/sea-legalcontext.htm>) states as follows:

- ▶ The SEA procedure can be summarized as follows: an environmental report is prepared in which the likely significant effects on the environment and the reasonable alternatives of the proposed plan or programme are identified. The public and the environmental authorities are informed and consulted on the draft plan or programme and the environmental report prepared. As regards plans and programmes which are likely to have significant effects on the environment in another Member State, the Member State in whose territory the plan or programme is being prepared must consult the other Member State(s). On this issue the SEA Directive follows the general approach taken by the *SEA Protocol* to the UN ECE Convention on Environmental Impact Assessment in a Transboundary Context. (Convention on Environmental Impact Assessment in a Transboundary Context done at Espoo (Finland), on February 1991; <http://live.unece.org/fileadmin/DAM/env/eia/documents/legaltexts/conventiontextenglish.pdf>)

Decision-making and participation processes are often studied and carried out under the broad umbrella conception of *governance* in the research field (European Commission 2001; OECD 2003; Pierre 2000; Renn 2008). The development in the field clearly shows the necessity of multidisciplinary and multi-stakeholder approaches for capturing the wide scope of important perspectives, as well as the detailed accounts of risk matters. The basic assumptions and general approach presented above in relation to the RISCOM model should be seen in that wider context. ➤ *Figure 29.2* gives examples of types of forums used in discussions of various environmental risk issues. An example from the Select Committee on Science and Technology in the United Kingdom, House of Lords (2000) clearly states the preferred future development of public dialogue:

- ▶ *That direct dialogue with the public* should move from being an optional add-on to science-based policy-making and to the activities of research organizations and learned institutions, and should become a normal and integral part of the process (paragraph 5.48) House of Lords (2000).

The following part of this background will give a short account of the methodological approach of small group discussions, i.e., techniques including, e.g., focus groups, dialogue

Forms for participation	Only experts	Individual "stakeholders" & decision-makers	The general public	Public decision-makers
Informative		Dialogue	Science shop	
Advisory	Science court Expert committee	Team synergy	Consensus conference	Strategic Environmental Assessment (SEA) EIA forum

Fig. 29.2

Examples of various forms of participation, types of groups involved, and aim of participation. Based on Andersson (2001), in the context of nuclear waste management. For elaboration see Andersson (2008)

groups, and group interviews. The reason to exclusively deliberate on group discussions, although there are many methodologies that can be and are used in risk communication work, is that each project is unique to a high extent, and it is usually necessary to start with getting an overview of what ideas, perceptions, and actions that are involved. Pilot projects are therefore often designed to extract as much qualitative information as possible from the involved participants using interviews and group discussions. Results from such projects can later on be utilized in new and more structured data collections that allow for many participants, e.g., questionnaires. Large projects can also deal exclusively with information collected from various kinds of interviews and group discussions.

Various forms of group discussions are exemplary ways to get personal views and collective reflections from participants in a project. Group discussions have the advantage of saving time as compared to personal interviews with the same number of individuals. A group discussion is different from a face-to-face encounter since several persons interact at the same time and thus influence each other. It is therefore neither an optimal choice of methodology if the goal is to investigate, e.g., individuals' attitudes or personal values, nor an ideal forum for exchanging thoughts of an intimate personal nature or for extremely timid persons to express their views. Group discussions are effective, however, for achieving a broad overview of an issue by the sampling of comments and, e.g., information needs. The reflecting and commenting on thoughts introduced by others in the group usually provide additional information in comparison to several isolated interviews. The research literature presents much scholarly material and insight of how to structure discussions and how to evaluate results (Merton 1987; Glaser 1992; Denzin and Lincoln 2005). Only a skeleton version is presented here to give a flavor of goals, work considerations, and requirements.

For research, and for information purposes, the goal of the group discussion often involves eliciting comments, ideas, and perspectives from several participants at the same time. Focus groups can take many forms (see, e.g., Kamberelis and Dimitriadis 2005). Groups can be “homogeneous” or “heterogeneous” in composition. For example, the design of a study could differentiate discussions of youth and adult groups, use separate groups of experts and non-experts, or have as a goal to involve representatives of a variety of interests in the same discussion. The choice of how to organize the composition of the groups involved in a project depends on the aim of the study, the type of issue discussed, and, e.g., actual or potential conflicts between groups or persons. Sometimes it can be helpful to start out with separate homogeneous groups to make sure that each stakeholder’s interest or perspective is fully investigated and captured before moving on to a second round of group discussions involving representatives from various perspectives to develop the discussion.

Preparation is essential! Before the actual meeting, there have been considerations of the “design” of the group(s) and discussion themes, the preferable time and place, what is enough time for a good result, and how to follow up. There are also a number of practical questions that need to be taken care of, e.g., who is the host, who pays what (e.g., about invitations and travel, localities, refreshments, etc.), and what information should be provided to the participants (e.g., financier, responsible organizer, person in charge, time, and locality) so they have access to a correctly presented situation when they decide on whether or not to be involved. Participants have been contacted, e.g., by letter and phone, and given information, descriptions of practical arrangements, topics, and types of questions.

A “good” or effective group usually involves five to eight persons. Note that the more persons involved, the more time must be allocated for the discussion. A group process takes time and due consideration must be taken to create a comfortable situation where project information, presentations of the participants, and time for their questions on the project and proceedings are parts of the starting up phase. Introductory information also describes the project, ethical rules in research, how the meeting will be structured, the future use of the information, and the expected use of the results. It is important to have clearly formulated themes and questions. Written material clarifying such matters may have been sent to participants in advance.

The degree of structure in the actual work, and the interaction rules, must be outlined and found in accordance with the aim and design of a project. The design of the group discussion usually involves one or a few central topics. The group leader distributes the turn taking in the discussion. And although the term “discussion” involves an open exchange format it is nevertheless important to impose a certain structure on the discussion for time reasons. Usually it is favorable to discuss one topic exhaustively before moving to the next. “Exhaustively” means that each participant has had the possibility to present his or her view without interruption, using reasonable time. Comments and reflections related to the views presented by other participants can be welcomed and encouraged if time allows, but such additions must aim at contributing new substance, not at evaluating others’ views. The session is *not* about establishing the “right” or “wrong” opinions. It is an exercise of acquiring information on the participants’ views with respect to a specific topic, with as much explanatory argumentation as possible.

The group discussion leader must be well prepared to follow up and ask for clarifications during the interaction. It is also common that the leader of the discussion summarizes the main results of the work before asking the participants for final comments. The time allocation and

degree of structure of this feedback depends on the design of the project. The information collected from a group discussion should also be summarized in writing in a way appropriate to the project's aim. The presentation format of such summaries can vary between verbatim transcriptions of everything uttered in the group to structured overviews of central ideas or responses. See the research literature for more detailed accounts of analysis and presentations of qualitative data (e.g., Denzin and Lincoln 2005). A written summary or report in the native language of the participants is usually expected in this kind of project. The publication of articles in scientific journals makes the results available to the larger research community for criticism and inspiration, in accordance with the idea of cumulative development of knowledge. The results from a project will in such a manner contribute directly and indirectly to the research field, and be available when considering new tasks and projects.

## Current Research

---

This section presents examples of social challenges, group interactions, and basic problem dimensions in selected research projects. It aims at highlighting the complexity of the situations where risk communication is introduced, as well as at exemplifying risk communication strategies and tools in fieldwork. Most examples illustrate work in local communities where reactions among the citizens were important for the initiation of the risk communication projects. One example emanates from an international European Union project that is included to highlight similarities and differences in relation to the same type of task across countries.

### Social Challenges: An Example of Public Reactions to Inviting a Geological Investigation

---

In an attempt to carry out geological investigations for a repository for spent nuclear fuel, the operator, Swedish Nuclear Fuel and Waste Management Co (SKB), reached out to Swedish municipalities in the beginning of the 1990s and suggested a dialogue on the issue. The geological investigation aimed at determining whether or not the bedrock conditions were suitable for further industrial explorations. The company received the first invitation from the municipality of Storuman in the north of the country. When the invitation became publicly known it quickly became a major political issue eventually resulting in a local referendum. The result of the referendum was a clear vote against the geological investigation (72%) and the invitation to the operator was withdrawn. The municipal board's initiative caused an unexpected social turbulence in the vast and sparsely populated northern municipality. It was of interest to the company to investigate the reasoning behind the outcome of the referendum in more detail and they therefore initiated a project (Drottz-Sjöberg 1996, 1999). The results were based on interviews with people who had been active in the campaign preceding the referendum. The participants were initially grouped into the two major interest categories of voting for or against the site investigation. Those who voted against a geological investigation emphasized values such as traditions, personal control, preference for small-scale business and decentralization, as well as keeping nature untouched, and they did not want to be dependent on large entrepreneurs. Those who voted in favor of inviting the operator instead

emphasized the need for major investments in the local area and jobs for people, economic expansion, and increased centralization.

However, the results also indicated influences from a major value dimension, i.e., “collectivism-solidarity vs. individualism-personal success.” This dimension split the main pro and con groups into subgroups. Thus, in the group promoting the geological investigation the municipal authority, with many years of firmly rooted social democratic solidarity values and the collective interest in mind, was found on the same side as the established business and industry part of the society. Those who had voted against in the referendum formed two major subgroups as well; one which especially emphasized traditions, decentralization, solidarity and the safekeeping of an intact wilderness, and one of small-scale business and potential entrepreneurs that disliked dependence on large entrepreneurs, but emphasized small family controlled businesses and making use of nature in a restricted way for personal gain.

Thus, the political initiative to invite the geological investigation, based on evaluations of the future common good of the municipality, found itself on the same side as the established industry in the referendum process. Small private entrepreneurs, afraid that their business offers, advertising clean and untouched nature, would be tainted by images of radioactive hazard and influential large-scale industry, came to side with strong local traditional values and village community groups. One lesson learned from this work was that there are various layers of values guiding opinion and action, such as deeply felt personal values and values guiding more strategic goal attainment. Another lesson was that a process that is not solidly anchored in the local population before a strategic decision is taken may quickly get out of hand, politically and socially, increase group polarization, sustain apprehensions of hidden agendas and block discussions and dialogue.

## **Group Challenges: An Example of Concerns of Politicians and Nuclear Experts**

---

The following example emanates from the “Communication 2000” project initiated by the Nordic Nuclear Safety Research (NKS) framework (Andersson et al. 2002). It aimed at studying problems in risk communication involving different professional groups. The project idea was triggered by a media event in 1998 when results from a probabilistic safety analysis (PSA) related to a Swedish nuclear power plant (O2) in the south-eastern part of the country was unfavorably compared to PSA results of a Lithuanian nuclear power plant. The background includes a standard report from the former plant to the Swedish Nuclear Power Inspectorate (SKI), written by experts for the same type of experts, which was publicly available in accordance with Swedish law. A freelance journalist compared these PSA results to PSA results available from the Ignalina nuclear power plant, and the resulting headlines announced that the risk of an accident was significantly higher in the Swedish plant.

The headlines and articles following this comparison caused concern at the national level, and especially in the O2-reactor’s home community of Oskarshamn. Politicians of the local government, and specifically the members of the local safety council responsible for the safety in the municipality, were suddenly, rather forcefully and very early in the morning faced with complex technical questions about PSA results, risk comparisons, and the quality of their overview work.

However, the turmoil was based on a misunderstanding, i.e., the assumption that PSA analyses from different plants could be compared. This is not the case since such analyses are uniquely related to each specific construction. The media event, subsequently denoted a “nonevent” because it lacked substance, nevertheless had its social effects and stimulated further interest in communication problems. The “Communication 2000” project aimed especially at identifying the circumstances, factors, and issues that after the media event were seen as essential in problematic information and communication situations related to nuclear safety. The steering principle of the project was based on the RISCOM idea to enhance transparency (Andersson et al. 1998, 1999).

Focus groups and questionnaires were used to identify and measure the problems encountered by persons involved in the situation. A pilot study engaged two focus groups mixed with respect to participants’ background, i.e., personnel at the nuclear power plant and politicians of the local safety board. The subsequent empirical data collection utilized central categories of problems elicited from the focus group discussions, and involved similar groups of nuclear power plant personnel, politicians, and administrators in the local community as respondents. The results showed, e.g., that both groups found it problematic to understand the way media works. Furthermore, nuclear power plant personnel highlighted the difficulty of explaining a subject matter in their field in a comprehensive way, whereas politicians and administrators were more concerned about the distribution of information. A significant difference was that the group of politicians and administrators rated information leaks as more problematic than the other group did. The theme “Problems in information transmission” could be used to construct two subindices measuring “structural problems” and “human problems” in information transmission respectively (For details, see Drott-Sjöberg 2001).

Situations that were perceived as especially difficult to handle included to explain something in front of national TV cameras, and to find the time to thoroughly study a subject matter. The politicians found it more difficult to understand mathematical formulas and expressions, and the nuclear power personnel found it more difficult than the politicians to present selected issues in front of a larger group of the general public. Five subindices could be constructed on the basis of responses to the theme “Handling situations” and these were “communication ability,” “competence,” “ability to synthesize information,” “context uncertainty,” and “bridging ability” (i.e., to evaluate subject material outside the own profession, to discuss with colleagues who usually have a different view, and to be confronted with aggressive persons at a meeting). Finally, the study pointed out three areas for improvement of the communication situation: to enhance understanding and clarity, to work with relations and contexts, and to intensify professional development and feedback.

Thus, this project revealed several aspects of the problematic side of communicating in general and about communicating risk in particular. It was noteworthy that the two groups highlighted different aspects of problems related to information presentation, e.g., nuclear plant personnel mentioned problems of presenting complex issues from their field of expertise in an easily accessible way and the politicians mentioned problems connected with comprehending core subject matters, as well as inappropriate, or strategic, information leakage. The discussions in the mixed focus groups were important for many reasons. For example, they improved the personal knowledge of other individuals, their roles and tasks. The discussions dwelled on what information was the most essential for various actors, how to read and especially how to write a technical report so that nonexperts could make use of it, etc. The various theoretical categories or indices mentioned above outlined problematic

communication aspects in situational, structural, interpersonal, and personal/professional areas. The specification of such sub-domains suggests that communication is relevant on many levels and that managing a specific risk situation is related to the identification of its unique challenges as well as the ability of the involved professional groups to communicate in all these areas.

## Challenges of Contents of Risk Information: An Example of Localization of a LNG Storage Tank

---

This example highlights risk communication work in relation to reactions to the localization of a liquefied natural gas (LNG) storage tank in Risavika, Sola municipality, neighboring the city of Stavanger, Norway (Vatn et al. 2008). Natural gas from the North Sea is transported through pipelines to shore, and then liquefied at a process plant before storage in a huge tank. The natural gas is then distributed from the facility to local consumers by LNG ships and LNG lorries. The initial approval process included two steps, firstly the facility plan was approved according to the development plan by Sola municipal in June 2006, and secondly were the technical aspects of the facility approved by central authorities according to the fire and explosion law in Norway in December 2007.

The planned localization nevertheless gave rise to an active debate in the surrounding municipalities via the media, as well as among neighboring residents. According to the SEVESO II Directive, the involved company shall develop a strategy for risk communication and information related to the emergency preparedness measures established. The relevant company did therefore arrange and invite to several public meetings as a part of this strategy. However, several factors contributed to an intense debate in this context. One aspect was that the company had claimed that the facility did not represent any risk to third parties at one of the public meetings, another that substantial critique was raised by profiled safety researchers in the region. Early risk analyses had shown a very low risk, but disclaiming the risk resulted in a serious mistrust of the company. The media attention, including a series of articles and letters to the editor in the local press, indicated that the risk communication process hardly was on the right track. In this context, and based on previous studies related to societal security and risk communication, the organizations of SINTEF/NTNU in Trondheim were approached by the company and asked to jointly assist in a risk communication process as an independent subcontractor.

The risk communication project was conducted in 2008. It aimed at explicitly identifying the risk issues that were considered the most demanding in various groups, and also to identify aspects of relevance to future risk analyses. The work involved the following steps:

- Explicit discussions with the funding company about work methods to be used, and about research independence regarding the carrying out and reporting of the work, before a contract was signed. For example, the report should summarize and present contributions in such a way that identification of individuals was not possible, and the raw data material was to be available only to the researchers.
- The company invited to a public meeting through announcements in local media. The meeting was held in the evening to allow participation for as many interested persons as possible. The chosen locality was situated in a school building to enhance a neutral ground for the meeting.

- The public meeting involved presentations by the company of the current situation and plans, by the research team on planned risk communication work, as well as unlimited time for questions and answers. The public meeting facilitated the identification of key interests. Individuals were approached for participation in focus group discussions later in the process.
- Further elaboration of key interests and central types of stakeholders, and on how focus groups should be designed to facilitate a thorough data collection and discussion. Four groups were eventually formed, i.e., groups of *risk analysts*, *neighbors*, i.e., individuals who lived in the vicinity of the harbor and the planned facility, persons working in *neighboring companies* localized in the harbor area, and personnel from the *fire brigade*.
- Focus group discussions were conducted separately in each group at times convenient to the participants to enable an in-depth penetration of all aspects considered important from their perspective (2–4 h per group).
- Summaries, including citations from the discussions, were sent to all participants of each relevant group for their comments and clarifications. These summaries constituted the basic materials of the final report on communication.
- There were two reports from the project, i.e., one report focusing on risk analytic aspects and one focusing on the risk communication process (Vatn 2009; Drottz-Sjöberg 2008). These were sent to all participants, and later made available on the Internet.
- A public meeting was held to inform about the results, and invite comments.

The results of the risk communication process showed that all groups wanted more information, including additional risk and safety analyses of the planned industrial site and the related transports, and on emergency planning. Concerns for “third parties” were central, e.g., potential accidents involving the neighboring area of housing and businesses. The participants had questions and comments related to the preceding decision-making process and to the status of local preparedness and rescue efforts. Other concerns involved the possibility of a future expansion of the planned capacity of the plant, the safety of regular passenger traffic, and leisure activities in the harbor area.

Some comments and requests from the focus group discussions related to the risk analysis. These included suggestions to provide more scenarios in the analyses, including “worst-case” scenarios, to better specify uncertainties, to improve the dispersion analyses (which were considered incomplete), to investigate possibilities of “domino effects,” i.e., sequences of negative events, and to consider effects of lightning, sabotage, and terrorism. The participants wanted explicit comparisons to international standards, and asked for inclusion of comments showing how previous mistakes or faults in similar facilities were followed up. They also asked for descriptions of what happens in the LNG-process, evaluations of possible problems in the process, including safety distances and potential hazards related to transports by sea and land. Maintenance aspects were brought up together with safety culture issues, training, and preparedness plans. Long-term health consequences were an issue as were possible effects of minor accidents. It was stated that, of course, it is the most terrible accident that is the most feared, but that also fear of losing competency in the own or surrounding businesses due to uncertainties attached to the risk issue is of importance, as is the inability to provide employees with information about risks to health and safety.

In addition to the mentioned aspects of expected risk information contents, there were comments on the information availability and decision process, especially related to the time

preceding the initiation of the LNG tank construction. The participants pointed out that 4,000–5,000 persons were in jobs in the area and that 6,000–7,000 persons worked or lived within a 3 km radius. Many felt unsafe and wanted more specific information, which they had not received so far. They criticized that the early risk analysis report was not officially available, and asked about the reasons for not informing thoroughly the neighboring businesses and the inhabitants in the residential area. Some perceived the entrepreneur as a “strong actor” with respect to resources, but a party that nevertheless had acted more unresponsive or reactive than proactive and forthcoming with respect to various concerns. The lack of information, unavailability of the preliminary risk analysis, and a decision process involving several municipalities lacking transparency put the trustworthiness of several actors into question. In short, the results showed discontentment with the availability, amount, and quality of the early risk information as well as with the transparency of the decision process. A number of scenarios and additional considerations were suggested to be included in the future risk analytical work, and the need for increased transparency of the decision-making process was highlighted. (For details see Drott-Sjöberg 2008; Vatn 2009.)

## Challenges of Contents of Risk Information: An Example of Safety Data Sheets

Risk communication is a major risk management strategy, and involves also written information. The project summarized here was part of a larger program called Achieving GREater Environmental Efficiency (AGREE), based at the Royal Institute of Technology, in Stockholm, Sweden, with funding from the Swedish Environmental Protection Agency. We investigated professional users' familiarity and views of safety data sheets, how they understood and evaluated the information presented, and their need for additional information.

Safety data sheets are provided by manufacturers of chemical products to professional users and include health, environment, and safety information. Such information may, e.g., include statements like: *the substance may cause irritation to eyes, skin, and respiratory tract. Repeated contact with the skin may cause dermatitis in sensitive individuals. Handling of this product may be hazardous.* What does this mean?

A pilot study asked a number of participants to give their personal interpretations of central concepts widely used in the field, e.g., “ordinary caution,” “chronic effect,” and “dangerous (classified) chemical.” We also wanted to investigate effects of inclusion or omission of specific information in the safety data sheets. In the main study, 70 individuals in professions handling chemicals on a frequent basis (personnel in the rescue services, the chemical industry, and safety personnel) participated in the study (Drott-Sjöberg and Drott 2004). The results showed that professional users were well aware that the handling of dangerous chemicals poses risks, and that experience of risk and uncertainty were more pronounced for long-term health effects. However, we also found quite large knowledge variability in professional user groups, and they understood selected expressions and concepts in a variety of ways. For example, the term “carcinogenic” was understood in two major ways. Most respondents explained in their own words that such a substance “might cause cancer” whereas a minority believed that it meant that the substance “causes cancer” – thus a difference between a risk of a health effect and an actual health effect when handling the substance. Some gave more unspecific answers, such as that the substance is dangerous and should be handled with

outmost care. The understanding of the concept “classified” was especially problematic. When asked about the interpretation of the expression “the substance is not classified” a majority believed it meant that the substance had not been investigated or tested, others thought that it was a warning, or that it was not dangerous, and still others that the substance was not approved by authorities.

The results also showed that if specific information on health or environmental effects was not provided in a safety data sheet the omission caused uncertainty, specifically among personnel with more experience with chemicals. However, a fair amount of the participants interpreted such omissions as indicating that, e.g., health or environmental effects were irrelevant to the context since they were not mentioned. Omissions of specific information generally resulted in lower ratings of the importance of such information for the handling of dangerous chemicals. Overall the study identified problems related to the information given in safety data sheets, e.g., lack of detailed and easily interpreted information, the ambiguity inherent in certain concepts, or in abbreviated information presentations. Thus, the results underline the importance of clarifying written information, e.g., ambiguous or technical concepts, and to test written information to better meet users’ information needs. The safety data sheets provide a specific type of challenge since useful information preferable is crystal clear as well as brief. And although these information leaflets are produced mainly for professional users the sheer number of chemical substances and products, as well as the variation in background knowledge, indicate that there are good reasons to continue the work to enhance information clarity. The recommendations from the project included to explicitly explain the intended meaning of terms and concepts, to provide better descriptions and characteristics of the attached risks, and their proper handling in normal use as well as in an emergency situation.

## International Challenges: Examples from the ARGONA Project

Four years of collaborative work, including a risk communication part (WP4), within the project Arenas for Risk Governance (ARGONA) resulted in more detailed knowledge and understanding of different countries’ nuclear waste management (NWM) strategies and achievements (Päiviö Jonsson et al. 2010). The countries investigated more closely in the risk communication work package were the UK, the Slovak Republic, the Czech Republic, and Sweden. The countries provided an interesting diversity with respect to historic developments, and strategies for information and communication on risk and safety aspects of nuclear waste management, as well as regarding organizational and governance structures. A few comments on similarities and differences across the countries are mentioned here, together with trust issues, the challenge of keeping a communication process alive, and lessons learnt with respect to strategies and tools in risk communication.

The intention of WP4 was to delineate good risk communication approaches across national borders, as well as to identify circumstances that require more specific considerations. The international workshop in the last project year, aimed at an in-depth discussion of risk communication related to management of nuclear wastes based on experiences from various stakeholders from the involved countries (Drottz-Sjöberg et al. 2009), as well as to go more deeply into communication of quantitative estimates related to risk and uncertainty (Bolado 2009). Only the former task is exemplified here.

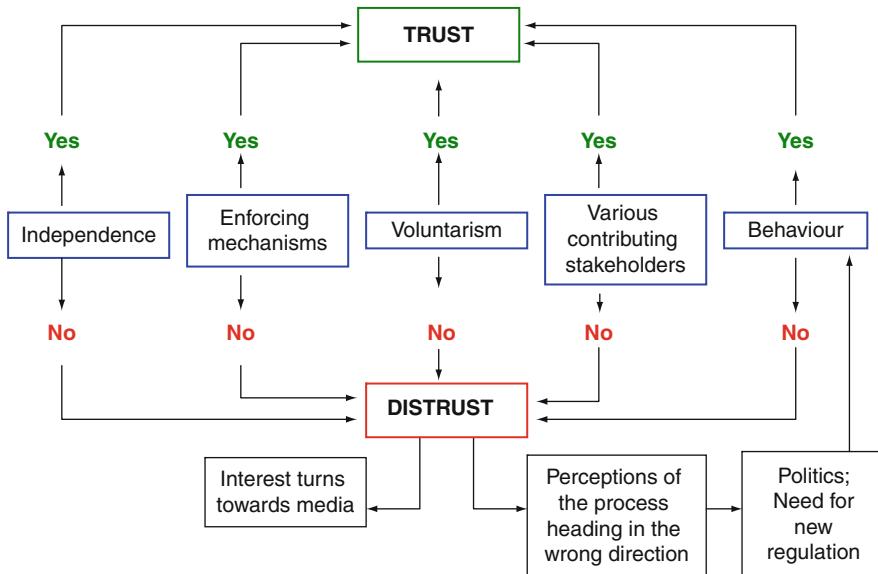
A number of stakeholders from the participating countries had the task to provide comments on strategies for risk communication that have emanated from various interest groups in different countries. Their engagement also aimed at eliciting comments on strengths and weaknesses associated with various risk communication techniques and more composite risk communication strategies. The work performed in the project fits well into the transparency framework provided by the RISCOM model, successfully used in Sweden and later applied in the Czech Republic within the ARGONA project.

Similarities of the participating countries included that they all are democratic European Union countries, and that NWM is a national issue and responsibility. International standards are guiding the work, including the use of Environmental Impact Assessments (EIA), and the countries collaborate with IAEA. There is also the similarity of awareness of the importance of information to a larger public, and that nuclear waste policies and management attract media attention. All participating countries have high qualification requirements of experts working in the field and the general population has an overall high educational level.

There are several differences between the countries as well, including what type of wastes that must be managed, the economic situation, the availability and status of funds for financing repositories, and the governance systems of NMW management. Differences also relate to historic events that influence current attitudes, levels of trust in various actors, the structure and functioning of the overall social system, e.g., with respect to openness, transparency, and traditions of communication, and last but not least, expected developments in the foreseeable future, especially if new nuclear power plants are considered to be built or not.

The discussions very quickly focused on the importance of trust. An exchange of views took place on the differences in trust in various countries and possible reasons for the differences. It was considered how much authority and official bodies are trusted, and to what degree such organizations are seen as *independent* from an implementer. It was noted, e.g., that historical events and decisions have affected trust in authorities. In some countries there seemed to be a view that “independence” involves not only acting separately vis-à-vis an implementer but also not being related to the government. In contrast, municipalities involved in the siting process in Sweden had chosen to use state authorities as “their experts” for addressing long-term safety issues and safety assessments. Regarding the UK, however, it was stated that much of what emanated from the authorities was perceived as “smoke and mirrors,” and that people instead tended to turn to the media for information. It was seen as hard to know and to understand what agendas the different actors pursued. A comment related to the Slovakian experience pointed out that trust actually may lead in the wrong direction and that certain skepticism is necessary. There was agreement that for trust to develop there is a need for *enforcing mechanisms* to follow up on mistakes or bad conduct irrespective of actor.

A comparison between the UK and Canadian approaches was made, and a proactive local risk communication approach was suggested as being more productive than situations where nuclear waste management organizations tour municipalities to feed information into a process. It was added that there is a danger in “forced risk communication” with respect to achieving trust. An example of how trust is enhanced in an unobtrusive way was provided from the UK, and the Seascate area in particular, where people working at Sellafield subsequently have chosen to retire in the coastal village. Their choice of staying close to the plant and industrial area was seen to provide an unspoken example of their attitude. Current activities in the UK involve *volunteers* (municipalities), and it was suggested that progress will be made due to development of a “partnership” approach in the UK (in West Cumbria).

**Fig. 29.3**

A summary of key factors in the discussion related to judgments of trust and distrust

However, it was also noted that the NWM issue can be politically very unpleasant. Examples to support this notion can be found in several countries. Sentiments and examples of unpleasant events and conflicts were related to historical contexts and cultural settings, and differences between generations were discussed. It was suggested that the extent to which various "generational cultures" exist across and within countries, e.g., with respect to interests, concerns, or environmental attitudes, influence overall strategies and trust levels.

The composition of participants influencing a decision process, in terms of what type of interests or "stakeholders" that are involved, should be considered. The issue initiated a discussion on the importance of *representative* participation. For example, it was suggested that an attempt to involve environmental NGOs in the European Nuclear Forum failed because the former groups held that the Forum consisted of 90% industry-related individuals, resulting in a very strong focus on industry issues, which some NGOs felt unable to support.

The figure below summarizes the key factors mentioned in the discussion, and suggests that *independence* of actors, functioning *control mechanisms*, *voluntarism*, and the involvement of a *variety of stakeholders* and what they *represent* can influence the balance of trust and distrust. In addition, the behavior and actions of participants also influence the process. The figure illustrates that all the central aspects influence the process by strengthening either "trust" or "distrust." It must be noted, however, that a "trust" or "distrust" outcome was not seen as good or bad *per se*, but as a judgment within a complex situation also related to, e.g., goals and achievements (Fig. 29.3).

It was noted in the workshop that *trust* is imperative and that *voluntarism* is a necessary basis for risk communication and participation processes. However, it is important to ask the

question “Risk communication about what?” What is the *subject matter* and what is the *goal* of the communication? Is it about safety issues, specific analyses, information about plans, or decision processes, etc.? Again the importance of clarity of goals as well as contents was underlined and it was suggested to make use of, and to improve, already available materials of good examples from work in different countries. When discussing *trust* it is also important to differentiate between various kinds of trust, e.g., trust in information regarding what science claims to contribute or solve (epistemic trust), and social trust, e.g., in authorities or institutions (Sjöberg and Wester-Herber 2008). Issues related to different kinds of trust may elicit different kinds of responses.

With respect to knowledge and interest in knowledge acquisition, it is valuable to distinguish between local communities and a larger or the national population. Local groups and people living in a community tend to be better informed of local circumstances and events, their history and role, and there is more at stake for them than for others. Such knowledge is especially relevant to consider when risks can be geographically determined, and it is linked to interest and attitudes (see, e.g., Sjöberg 2008; Kari et al. 2010). The result of trying to get a message across to a group with little interest or no experience of the matter at hand would to a higher extent depend on generalized views, social trust, and credibility factors attached to the communicator. The type and context of a hazard or risk should therefore be taken into account in the information or communication situation.

The figure below aims at illustrating some central concepts of importance for keeping a risk communication process alive, and arrows pointing to the center support a continuation of the process. If a prerequisite for communication is not fulfilled, that aspect will contribute to the moving away from the core goal, i.e., drift into the surrounding area of “no-participation.” The arrows between the central concepts do not suggest causal flows but aim at illustrating that prerequisites for communication are interlinked and continuously influencing each other. It is also suggested that each central concept is available as a starting point for an improved communication process, as well as an end point if expectations are not met. Note, also, that if the goal of a communication process foremost is to keep the process sound and alive, then the maintained interaction itself represents the positive outcome (☞ Fig. 29.4).

Previous work in WP4 had shown that there exists a vast amount of information with respect to risk communication processes, and that it is important to continue to enhance democratic governance processes to develop the work. The project also found that there is a need to inform more precisely with respect to details, account for various levels of knowledge and involvement, and that a variety of approaches can be used in such work. It was suggested that interconnecting knowledge and experiences from different risk management areas and practices regarding threats to health, safety, and the environment may offer fresh perspectives. Broader approaches could help increase “the tool box” useful to risk communication work and also help to test what approaches fit best in diverse subject areas and contexts. The main conclusion from the discussion was: *Think European but pay attention to local detail*. Regarding the question if it is reasonable to compare risk communication processes across countries the conclusion was “yes” in relation to certain general approaches and communication tools, but “no” with respect to the feasibility of comparisons for approaching requirements of unique situations.

The table below structures some of the central input from the workshop in a different way. Five steps of a possible NWM process are outlined, together with some notions of content and requirements (☞ Table 29.1).

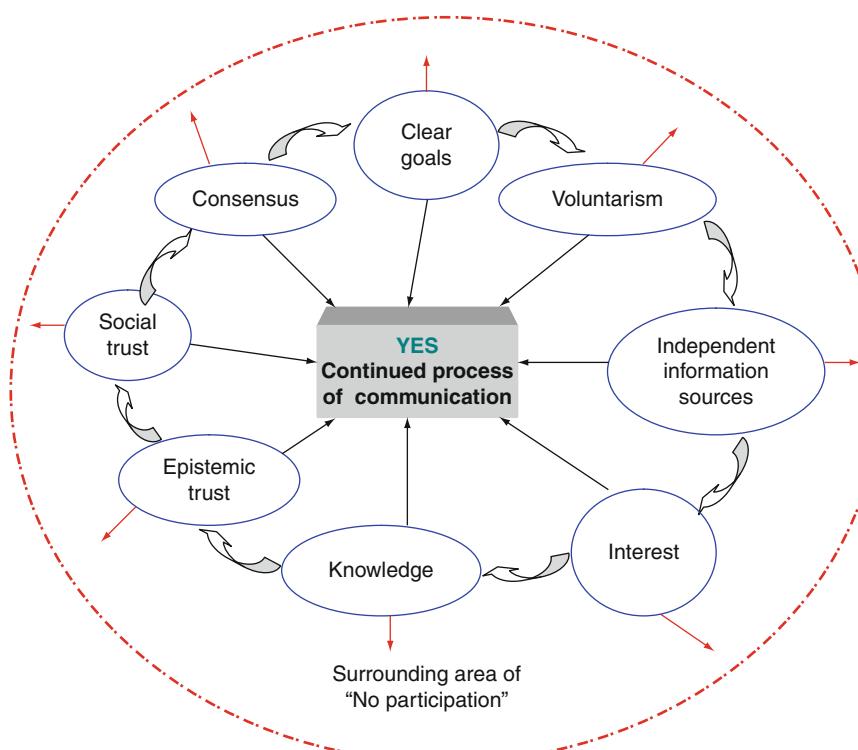


Fig. 29.4

A loop of prerequisites, and perpetual influences, in continuous communication processes

## Further Research

The examples provided above aimed at illustrating some of the reasons to initiate risk communication projects as well as aspects to consider in the work. The Storuman example showed, e.g., that a strategic political decision aiming at benefiting the public good did not gain public acceptance for many reasons, including that the decision to invite SKB to do a geological site investigation was not thoroughly discussed and broadly anchored in the municipality. It also showed that providing citizens with a dichotomous choice in a referendum effectively shadowed underlying considerations of importance for making that choice and instead intensified group polarization.

The “nonevent” media example highlights the social interconnectedness of professional groups carrying out very different functions, and points out the necessity of cross-disciplinary insights in tasks and functions to be able to perform well in the own professional field. It also shows that alarmist news can trigger very quick public responses and that media contents may affect daily life irrespective of the validity of the information.

The projects used to illustrate the importance of early and clear communication of risk information show, e.g., that there is a formidable source of perspectives, as well as expertise, outside of the more strictly defined boundaries of an industrial project, and that such

**Table 29.1**

**Overview of important major steps, and examples of contents and requirements, regarding a possible nuclear waste management process focused on implementation of a repository plan.**  
**Extracted from the ARGONA workshop 2009**

Steps	Content	Requirements
Strategic consideration	Provide reasons for planning a repository	Independence (reviewers, decision-makers)
	Clear definition of "safety case" and evaluation of feasibility	Transparency; independence; trust
	Choice of governance perspective of process	Functioning enforcing control mechanisms
	Clear criteria for regarding/disregarding a site	Consideration of "all" aspects
	Creation of inclusive social communication process	Social acceptance, openness
Agenda setting preparations	Clearly define the agendas for the technical and the social processes	Overview of uncertainties and clarification of "representativeness"
	Consider pace of development	Distinguish short- and long-term issues
	Proactive information	Information availability; understandable to target groups
	Prepare for involvement of stakeholders	Procedure perceived as fair
Contacts and discussions	Create interest	Preparatory work
	Invitation to participate	Voluntary processes
	Present uncertainties	Availability of pedagogical experts
	Discussions on, e.g., "the right community," "partnership"	Interested local communities
	Investigate possibilities of local steering mechanisms	Availability of local control mechanisms, e.g., veto-right
Recommendations on risk communication	Provide and keep available information in various formats	Clear goals of each intervention
	Use correct and understandable information materials	Invite to feedback on information materials
	Involve key group members	Extensive (local) network
	Collect and improve on available materials and experiences	Compilation of research results; new research
	Development of presentation techniques and skills; training in dialogue settings	Understanding of novel perspectives in the personal expertise area
	Work with "translations" of scientific terminology and ambiguous concepts	Cross-disciplinary collaboration; involvement of lay people
Implementation	Developments of concrete plans and work	Continuous updates of key factors related to technological and social developments

perspectives and expertise can be utilized to develop and refine analyses as well as to better meet information demands. The examples also show that it has consequences if information is presented or not. The LNG project involved social and psychological consequences of early information deficiencies enhancing distrust and worry. However, it is noteworthy that omissions of information in safety data sheets, e.g., regarding health or environmental aspects, instead seemed to lead users to conclude that omitted information was irrelevant to product use. The examples may suggest interesting research projects.

Is it the case that we ordinarily nourish a mental shortcut mode of thinking, i.e., rule of thumb or heuristic, which holds that if there is any danger at all someone will provide us with “the whole truth and nothing but the truth”? The existence of such a heuristic would help explain the current focus in the research area on the role of trust – an issue especially highlighted in the risk communication workshop of the ARGONA project mentioned in this chapter. Could it be argued that, in contrast to much current theorizing, we are not living in a “risk society” but in a “safety obsessed” society? Not because there are no hazards – there certainly are – but due to expectations based on unreflected beliefs that the world should be pure and safe and that any alteration of that state is due to external forces outside personal control and responsibility. Hence, *receiving* information from *trustworthy* sources becomes more vital than actively making sense of, or checking the validity of, information received. However, a person or a citizen cannot be reduced to filling only an information receiving function. The risk communication area provides ample examples of not only demands for information but the willingness – and sometimes the demand to be allowed – to provide input to a decision process. Thus, there seem to be possibilities to enhance and develop interactive communication, as well as personal control and management of risks, all in line with the wording of international conventions on participatory contributions.

Among the major challenges in the risk communication field lays the task of how to shift the attitude of information processing related to specific risks from a shortcut, heuristic thinking mode to deliberate cognitive processing. In accordance with dual-processing theories (e.g., Petty and Cacioppo 1986; Chaiken 2003) we become better decision-makers, and less vulnerable to unobtrusive influences, if we process information in a systematic and thorough way. Less cognitive effort due to use of heuristics is, of course, time efficient in everyday life, but there are no standard heuristics available that effectively manage risks. It seems as if trust has been tried as a surrogate heuristic and that its failure to meet expectations has led to a rather unproductive distrust discussion rather than a sharpened focus on information processing and risk mitigation efforts. If trust represents social capital, what does distrust signify? If distrust reflects decay of the social web, there ought to be intensified efforts to better understand the surrounding world for sheer survival reasons. Thus, to better prepare for encounters with hazards and for managing potential negative consequences we must participate, voice concerns, and present suggestions; as citizens we must collect information, scrutinize the validity of the information, discuss with others, and find ways to evaluate the identified risks and what needs to be done. Risk communication projects are one type of excellent arenas for communicating about risks and for acquiring knowledge as well as find links to additional information sources.

Openness, transparency, voluntariness, and participation are currently highlighted concepts in the risk communication and governance literature, as in society more generally. However, the use of the concepts in the literature sometimes involves more rhetoric and theorizing than information of concrete practical usefulness in risk communication work.

This state of affairs presents a challenge to research of following up on principles and theories by extracting their substance and systematically test it out in concrete settings. The importance of operationalization of concepts becomes very clear in such concrete work. For example, in these and other contexts it is useful to clarify that *openness* often refers to information access (and not some principle demanding exposure of inner secrets), *transparency* often aims at achieving clarity which enables understanding, e.g., of the structure of a task or framework, or the path and development over time of a decision process. *Voluntariness* is an important principle to respect in participatory processes and the concept highlights the challenges to motivate interest and to make interesting the subject matter that invites to participation for achieving a good solution.

Participation is discussed here from a somewhat different angle. In an essay in the booklet “Ethical and philosophical perspectives on the nuclear waste issue – eight essays by Sven Ove Hansson (2010)” the author addresses, among other issues, expert power, acceptance, voluntary risks, and steps in a decision process. In this context, it is of interest that Hansson argues that “risk decisions” cannot meaningfully be distinguished from other kinds of societal decisions and that almost all kinds of decisions in a society are to some extent “risk decisions.” He also notes, however, that in discussions on overall political decision-making the central concept is *democracy*, whereas in discussions on risk and risk decisions the central concept is *acceptance*. The former concept aims at decisions being in line with the view of the majority whereas *acceptance* focuses on what risks the general public, or those directly affected, can be made to accept. Hansson furthermore makes use of the French philosopher Condorcet, who in 1793 developed a suggestion to the French constitution to illustrate steps in the decision process. Condorcet outlined three steps in democratic decision processes: (1) Discussions of the general principles that will constitute the basis of the decision, investigation of the various aspects of the issues, and the consequences of different ways to take the decisions. (2) A second discussion where views are developed and combined with the purpose of generating a smaller number of general views. (3) The voting when the decision is taken. Hansson notes that, in relation to risk decisions, the discussion often focuses solely on the third step. The general public is presented as a group whose acceptance or agreement shall be obtained. He underlines that a reduction of the public’s influence to the third decision step regarding risk issues is inconsistent with the principles that generally apply in a democratic society.

Similarly, it was noted above that a citizen cannot be reduced to an information receiver. Is it possible to bridge the gap between science (including those with expert knowledge on risks) and society (including all citizens)? The European Commission (2009) has devoted some thought as well as research funding to this topic. The Directorate-General for Research, Science in Society set up the MASIS expert group to dwell on the issue and in 2009 they presented their report “Challenging Futures of Science in Society – Emerging trends and cutting-edge issues.” As can be noted they put science firmly *into* the society.

In the context of risks and communication, it could be noted that the communication concept indicates more input to an interaction process than one-way directed information messages. Risk communication involves the tasks to identify the risks, estimate them, and try to avoid them or manage them. The goal of risk communication is to eliminate or minimize the identified risks. It seems as if we live in a “risk society” with heightened risk awareness there may be less problems ahead than if we live in a “safety obsessed” society lacking trust in others. The emphasis in the research area, and in international guiding principles and conventions, on participation in decision processes could possibly point out a means of narrowing the gap

between passive information recipients and informed citizens, between science and society, and between acceptance and democracy.

The main point here, however, is that risk communication means to *communicate* about *risks*. This chapter has tried to highlight that the core of this type of communication is centered on the identification, estimation, and management of risks. In practice, however, there are several additional aspects to consider. Risk communication projects have, in addition, to managing situations ranging from doomsday beliefs to total disinterest, from inaccessibly rare expertise knowledge to pure ignorance, and from strong active stakeholder interests to deeply experienced injustice or victimization. These dimensions exemplify aspects that bring risk communication into the realm of the multifaceted and fascinating “real world.” And although the aspects are intimately attached to risk communication work, they must be kept at arm’s length from the central aims of identifying, estimating, and managing risks.

## References

---

- Aarhus Convention (1998) Convention on access to information, public participation in decision-making and access to justice in environmental matters, Aarhus. <http://www.unce.org/env/pp/documents/cep43e.pdf>. Accessed 25 June 1998
- Andersson K (2001) Mapping of processes using the RISCOM model. In: Andersson K, Lilja C (eds) Performance assessment, participative processes and value judgements. Report from the first RISCOM II workshop, SKI Report 01:52, Dec 2001. Swedish Nuclear Inspectorate (SKI), Stockholm
- Andersson K (2008) Transparency and accountability in science and politics. The awareness principle. Palgrave Macmillan, Chippenham
- Andersson K, Espejo R, Wene C-O (1998) Building channels for transparent risk assessment. SKI Report 98:5. Swedish Nuclear Inspectorate (SKI), Stockholm
- Andersson K, Balfors B, Schmidtbauer J, Sundqvist G (1999) Transparency and public participation in complex decision processes – prestudy for a decision research institute in Oskarshamn. TRITA-AMI Report 3068. Royal Institute of Technology, Stockholm
- Andersson K, Drottz-Sjöberg B-M, Lauridsen K, Wahlström B (2002) Nuclear safety in perspective. Final report of the nordic nuclear safety research project SOS-1. Report NKS-60. ISBN 87-7893-115-0. NKS Secretariat, Roskilde
- Andersson K, Drottz-Sjöberg B-M, Espejo R, Fleming PA, Wene C-O (2006) Models of transparency and accountability in the biotech age. Bull Sci Technol Soc 26:46–56
- Boholm Å (1998) Comparative studies of risk perception: a review of twenty years of research. J Risk Res 1:135–163
- Bolado R (2009) On the adequacy of the format proposed to communicate risk and uncertainty. ARGONA-report. Arenas for risk governance. FP6-036413. Deliverable D17. European Commission. Accessed 18 Dec 2009
- Breakwell GM (2007) The psychology of risk. Cambridge University Press, Cambridge
- Chaiken S (2003) Heuristics versus systematic information processing and the use of sources versus message cues in persuasion. In: Kruglanski AW, Higgins ET (eds) Social psychology. A general reader. Psychology Press, New York, pp 461–473
- Covello VT, McCallum DB, Pavlova MT (1989) Effective risk communication. The role of responsibility of government and nongovernmental organizations. Plenum, New York
- De Marchi B (1991) Public information about major accident hazards: legal requirements and practical implementation. Ind Crisis Q 5:239–251
- De Marchi B, Funtowicz S (1994) General guidelines for content of information to the public. Directive 82/501/EEC Annex VII, EUR Report 15946 EN, Office for Official Publications of the European Communities, Luxembourg
- Denzin NK, Lincoln YS (2005) The Sage handbook of qualitative research, 3rd edn. Sage, Thousand Oaks
- Drottz-Sjöberg B-M (1996) Stämningar i Storuman efter folkomröstningen om ett djupförvar [Sentiments in Storuman after the referendum on a deep level repository]. Projekt Rapport PR D-96-004. Swedish Nuclear Fuel and Waste Management Co (SKB), Stockholm
- Drottz-Sjöberg B-M (1999) Divergent views on a possible nuclear waste repository in the community: social aspects of decision making. In: Proceedings of the VALDOR conference, Stockholm,

- 13–17 June 1999. Karinta-Konsult, Stockholm, pp 363–369
- Drottz-Sjöberg B-M (2001) Communication in practice. In: Andersson K (ed) Proceedings VALDOR 2001, Swedish Nuclear Inspectorate (SKI), Stockholm, pp 419–427
- Drottz-Sjöberg B-M (2008) LNG-anlegget i Risavika. Kommentarer og synspunkter fra risikoanalytikere, bedriftsnaboer, brannvesen og beboere i Tananger {The LNG plant in Risavika. Comments and views from risk analysts, neighbouring companies, the fire brigade and residents of Tananger}. SINTEF, Trondheim
- Drottz-Sjöberg B-M, Drottz L (2004) Uncertainty and interpretations in risk communication. A study focusing on safety data sheets. Report in the AGREE-project. Royal Institute of Technology (KTH)/The Environmental Protection Agency (NV), Stockholm
- Drottz-Sjöberg B-M, Richardson P, Prítrský J (2009) Risk communication strategies. Conclusions and summaries of feed-back comments from participating countries. ARGONA-report. Arenas for risk governance. FP6-036413. Deliverable D18. European Commission. Accessed 13 Dec 2009
- Dunlap RE, Kraft ME, Rosa EA (eds) (1993) Public reactions to nuclear waste. Citizens' views of repository siting. Duke University Press, Durham
- EIA, Environmental Impact Assessment (1985) Council Directive 85/337/EEC of 27 June 1985 on the assessment of the effects of certain public and private projects on the environment. <http://europa.eu/legislation>. Accessed 25 Jun 2011
- European Commission (2001) European Governance. A white paper. COM(2001) 428 final, Brussels. [http://www.eur-lex.europa.eu/LexUriServ/site/en/com/2001/com2001\\_0428en01.pdf](http://www.eur-lex.europa.eu/LexUriServ/site/en/com/2001/com2001_0428en01.pdf). Accessed 25 July 2001
- European Commission (2009) Challenging futures of science in society – emerging trends and cutting-edge issues. Report of the MASIS expert group setup by the European Commission. Directorate-general for research, science in society. EUR 24039 EN. ISBN 978-92-79-72978-0. <http://ec.europa.eu/research/research-eu>. Accessed 25 Jun 2011
- Fischhoff B (1995) Risk perception and communication unplugged: twenty years of process. Risk Anal 15:137–145
- Fischhoff B, Bostrom A, Quadrel MJ (1993) Risk perception and communication. Annu Rev Public Health 14:183–203
- Glaser BG (1992) Basics of grounded theory analysis. Sociology Press, Mill Valley
- Gurabardhi Z, Gutteling JM, Kuttschreuter M (2005) An empirical analysis of communication flow, strategy and stakeholders' participation in the risk communication literature 1988–2000. J Risk Res 8:499–511
- Habermas J (1988) Om begreppet kommunikativ handling. In: Habermas J (ed) Kommunikativt handlande. Texter om språk, rationalitet och samhälle. Daidalos, Göteborg, pp 175–203
- Hansson SO (2010) Etiska och filosofiska perspektiv på kärnavfallsfrågan – åtta essäer av Sven-Ove Hansson {Ethical and philosophical perspectives on the nuclear waste issue – eight essays by Sven Ove Hansson}. Swedish Nuclear Fuel and Waste Management Co (SKB), Stockholm
- House of Lords (2000). Science and technology third report. <http://www.publications.parliament.uk/pa/Id199900/Idselect/Idsctech/38/3801.htm>. Accessed 25 Jun 2011
- Kamberelis G, Dimitriadis G (2005) Focus groups. Strategic articulations of pedagogy, politics, and inquiry. In: Denzin NK, Lincoln YS (eds) The Sage handbook of qualitative research, 3rd edn. Sage, Thousand Oaks, pp 887–907
- Kari M, Kojo M, Litmanen T (2010) Community divided. Adaptation and aversion towards the spent nuclear fuel repository in Eurajoki and its neighbouring municipalities. ISBN 978-951-39-4148-2. University of Jyväskylä, Jyväskylä
- Kennedy J, Delaney L, Hudson EM, McGloin A, Wall PG (2010) Public perceptions of the dioxin incident in Irish pork. J Risk Res 13:937–949
- Krauss RM, Fussell SR (1996) Social psychological models of interpersonal communication. In: Higgins ET, Kruglanski AW (eds) Social psychology. Handbook of basic principles. The Guilford Press, New York, pp 655–701
- Lasswell HD (1948) The structure and function of communication in society. In: Bryson L (ed) The communication of ideas: religion and civilization series. Harper & Row, New York, pp 37–51
- Levidow L, Carr S, Wield D (2000) Genetically modified crops in the European Union: regulatory conflicts as precautionary opportunities. J Risk Res 3(3):189–208
- Littlejohn SW (1999) Theories of human communication, 6th edn. Wadsworth, Belmont
- McQuail D (2000) McQuail's mass communication theory, 4th edn. Sage, London
- Merton R (1987) The focused group interview and focus groups: continuities and discontinuities. Public Opin Q 51:550–566
- Moffatt S, Hoeldke B, Pless-Mulloli T (2003) Local environmental concerns among communities in north-east England and south Hessen, Germany: the influence of proximity to industry. J Risk Res 6:125–144
- North W (1998) Comments on "Three decades of risk research". J Risk Res 1:73–76

- OECD (2003) Stakeholder involvement tools: criteria for choice and evaluation. Organisation for Economic Co-operation and Development, Paris, <http://www.nea.fr/html/rwm/docs/2003/rwm-fsc2003-10.pdf>
- Päiviö Jonsson J, Andersson K, Bolado R, Drott-Sjöberg B-M, Elam M, Kojo M, Meskens G, Prírský J, Richardson P, Sonerdy L, Steinerova L, Sundqvist G, Szerszynski B, Wene C-O, Vojtechova H (2010) Towards Implementation of transparency and participation in radioactive waste management programmes. ARGONA Final Summary Report. EU Contract FP6-036413. ARGONA Deliverable D23b. Accessed 12 Feb 2010
- Petty RE, Cacioppo JT (1986) The elaboration likelihood model of persuasion. In: Berkowitz L (ed) Advances in experimental social psychology. Academy Press, San Diego, pp 123–205
- Petty RE, Wegener DT (1998) Attitude change: multiple roles for persuasion variables. In: Gilbert DT, Fiske ST, Lindzey G (eds) The handbook of social psychology, vol I. McGraw-Hill, Boston, pp 323–390
- Pierre J (ed) (2000) Debating governance. Authority, steering and democracy. Oxford University Press, Oxford
- Plough A, Krimsky S (1987) The emergence of risk communication studies: social and political context. Sci Technol Hum Values 12(3–4):4–10
- Prades A, Esplugas J, Real M, Solá R (2009) The siting of a research centre on clean coal combustion and CO<sub>2</sub> capture in Spain: some notes on the relationship between trust and lack of public information. J Risk Res 12(5):709–723
- Renn O (1998) Three decades of risk research: accomplishments and new challenges. J Risk Res 1:49–71
- Renn O (2008) Risk governance. Coping with uncertainty in a complex world. Earthscan, London
- SEVESO II Directive. Council Directive 96/82/EC of 9 December 1996 on the control of major-accident hazards involving dangerous substances
- Sjöberg L (2008) Attityd till slutförvar av använt kärnbränsle. Struktur och orsaker {Attitudes to final repository for spent nuclear fuel. Structure and reasons}. SKB R-08-119. Swedish Nuclear Fuel and Waste Management Co (SKB), Stockholm
- Sjöberg L, Wester-Herber M (2008) Too much trust in (social) trust? The importance of epistemic concerns and perceived antagonism. Int J Global Environ Issues 30:30–44
- Stern PC, Fineberg HV (eds) (1996) Understanding risk: informing decisions in a democratic society. National Academy Press, Washington, DC
- Vatn J (2009) Risiko og beslutningsprosesser i forbindelse med LNG-anlegget i Risavika i Sola kommune {Risk and decision processes related to the LNG plant in Risavika in Sola municipality}. SINTEF Report A10107. ISBN 978-82-14-04715-8. SINTEF Teknologi og Samfunn, Trondheim
- Vatn J, Vatn GÅ, Drott-Sjöberg B-M (2008) Societal security – a case study related to an LNG facility. In: The Research Council of Norway Research programme “Societal security and risks – SAMRISK: is there a Nordic model for societal security and safety?” Paper in the proceedings of NFR’s conference, Oslo, 1–2 Sept 2008
- Watzlawick P, Beavin J, Jackson D (1967) Pragmatics of human communication: a study of interaction patterns, pathologies, and paradoxes. Norton, New York
- Wright D, Dressel K, Pfeifle G (2006) STAKEholders in risk communication (STAR). Good practices in risk communication. Deliverable 3. PRIORITY FP6-2003-SCIENCE-AND-SOCIETY-7



## Risk Ethics



# 30 Ethics and Risk

Douglas MacLean

University of North Carolina, Chapel Hill, NC, USA

<i>Introduction</i> .....	792
<i>History</i> .....	792
<i>Risk and Rights</i> .....	793
<i>Permitting Productive Risky Activities</i> .....	795
<i>Risk and Consent</i> .....	796
<i>Justice and the Distribution of Risk</i> .....	800
<i>Current and Further Research</i> .....	801
<i>Conclusion</i> .....	803

**Abstract:** Although risk is a fact of life, it was not extensively discussed in moral philosophy before the 1970s. Robert Nozick's classic discussion of risk drew philosophical attention to the special problems created by actions that create risk but which may or may not result in any harm. This chapter begins with a description and analysis of Nozick's argument. It concludes that familiar concepts in moral philosophy like harm, compensation, and individual moral rights cannot by themselves give a satisfactory analysis of the unique ethical problems created by activities that impose or attempt to regulate risk. This discussion leads to further examination of the relation between risk and consent. Appeals to consent are important in the justification of techniques of risk analysis used to reveal individual preferences for comparing risks, costs, and benefits in policy decisions. This discussion is followed by a review of issues involving justice and the distribution of risk. It focuses especially on some distributional issues that are unique to risk and reveal important differences between looking at the ethical dimensions of risk from an individual and from a societal perspective. The chapter concludes with some speculative remarks about future research into the ethics of risk that is prompted by increased awareness of risks imposed by new technologies, the prospect of reducing or mitigating the effects of anthropogenic climate change, and the increasingly likely prospect that decisions today may impose significant risks on future generations.

## Introduction

---

Until fairly recently, moral philosophers have largely ignored questions about the ethics of activities which create risk that may or may not result in actual harm. This is curious, since so many of our actions obviously involve taking risks and imposing risks on others. Nevertheless, the history of moral philosophy, at least in the modern era, has been predominantly concerned only with actual harm and the rights, obligations, and reasons for blame or compensation that arise with respect to benefit and harm. The subject of this chapter is the ethics of risk. We begin with a brief review of references to risk in the philosophical literature before the last quarter of the twentieth century. Then we examine the problems that have dominated philosophical discussion of the ethics of risk since that time, beginning with Robert Nozick's classic discussion of the particular ethical issues posed by risk. This is followed by a discussion of some of the leading issues involving risk, including consent, distributive justice, and issues raised by applying formal techniques of risk analysis in the context of public policy decisions.

## History

---

In the few instances in which philosophers writing before the last half century have mentioned risk, they have seldom paused to consider the special issues that it raises. This is not to suggest that risk has not been a traditional concern of moral and political philosophers. Indeed Hobbes's political theory can be described as aiming above all to remove the central risks of human life and provide the security that comes from entering into civil society (Hobbes 1660/1996). But neither Hobbes nor later moral and political philosophers pay much attention specifically to risk. For example, in his discussion of government action that permissibly restricts individual liberty, John Stuart Mill writes, "Whenever, in short, there is a definite

damage, or a definite risk of damage, either to an individual or to the public, the case is taken out of the province of liberty and placed in that of morality or law" (Mill 1859/1978). Clearly, on Mill's view, it is permissible to restrict a person's liberty to prevent harm to others, and those who are harmed through negligence may demand compensation or restitution. But what is a proper response for those who have been exposed only to the risk of harm? Mill's example of risky behavior gives us little help in answering this question. He writes: "[W]hen a person disables himself, by conduct purely self-regarding, from the performance of some definite duty incumbent on him to the public, he is guilty of a social offense. No person ought to be punished simply for being drunk; but a soldier or policeman should be punished for being drunk on duty" (Mill 1978, pp. 79–80). The actions of the drunken policeman increase risk to the public, of course, but the policeman is also in breach of a duty, which is itself a wrong. So the example tells us little about what special problems, if any, are caused by risk-imposing actions per se or by actions that impose risk but do not otherwise violate a duty or responsibility.

Henry Sidgwick says a bit more than Mill about the ethics of actions that impose risk and our different responses to those actions. Alluding to Mill, Sidgwick writes, "Again, certain practices dangerous to others – such as the drinking of alcohol to the point of intoxication – may be tolerated in private but repressed if the drunkard appears in public" (Sidgwick 1904, ch. 9, sec.2). But the perspicacious Sidgwick, who was a more thoroughgoing utilitarian than Mill, carries the argument further. He sees no problem with inhibiting individual liberty for the sake of reducing or managing risk of harm to a greater public, thus justifying "an extension of governmental interference, in the way of regulation." Sidgwick continues: "The easiest and most effective way of preventing harm is to prescribe certain *precautions* against it – i.e., to prohibit acts or omissions not directly necessarily mischievous to others, but attended with a certain risk of mischief." This statement is followed by a number of examples, ranging from government inspections for safety or requirements for standardized weights and measures that reduce the risk of fraud, to "restrictions on the manufacture and carriage of explosive substances and rules against importing cattle from countries where disease is rife. It is not certain that any given cargo of suspected cattle or carelessly carried explosives would do any harm: but most prudent persons see that the risk is too great to run" (Sidgwick 1904, ch. 9, sec. 2).

To philosophers whose views are more libertarian or rights based than Sidgwick, the idea that it is permissible to restrict or regulate activities simply in order to reduce risk, especially when the risks may never result in actual harm, is more problematic. And although many philosophers have found inspiration in Immanuel Kant's philosophical writings to develop a Kantian framework for thinking about the ethics of risk (Gillroy 2000), one cannot find any explicit discussion of risk in the writings of Kant or most other nonutilitarian philosophers before the 1970s.

## Risk and Rights

In contemporary philosophy, Robert Nozick was the first to discuss ethical issues involved in risk extensively (Nozick 1974, ch. 4). Acknowledging that "no natural-law theory has yet specified a precise line delimiting people's natural rights in risky situations," Nozick considers

three possible responses to actions that impose risk (Nozick 1974, pp. 75–76. For clarity, I substitute the phrase “rights violation” for his “boundary crossing”):

1. The action is prohibited and punishable, even if compensation is paid for any [rights violated], or if it turns out to have [not violated any rights].
2. The action is permitted provided compensation is paid to those persons whose [rights actually are violated].
3. The action is permitted provided compensation is paid to all those persons who undergo a risk of a [rights violation], whether or not it turns out that their [rights are actually violated].

(3) has both advantages and disadvantages compared to (2). If people are compensated for being exposed to a risk of harm, they may decide either to keep that gain or to pool their gains in order to provide greater compensation to those who are actually harmed. Thus, (3) gives people a greater degree of liberty to respond to risk imposition. But to compensate people for being exposed to *risk* is to treat the expectation of harm as itself a bad thing, which is puzzling. Someone who is under a risk or threat of harm may experience fear or anxiety, which is an evil, but it is not clear why the risk itself should be considered a kind of harm or evil that deserves to be compensated, especially if the actual harm never occurs. This is especially true if the person exposed to risk is unaware of this fact, and the risk never results in harm. These may be reasons for favoring (2) over (3).

But (2) has other problems. First, some of the risks that are imposed on people involve a risk of death, and if that harm occurs, then the victim cannot be compensated. Moreover, some of the risks we face are caused by the collective actions of many people in such a way that the action of any individual creates only a minuscule risk while the similar actions of many people create substantial risk, as in the case of standard kinds of “nonpoint source” air or water pollution. These risks are created by the collective actions of many people, and they fall on many others, sometimes at great distances from their source. The idea that those who suffer the harms from such risk-imposing activities – for example, those who contract cancer or some other disease as a result of exposure to nonpoint source pollution – can be compensated in a proportionate way by all the individuals who create the risk, is of course entirely impractical.

Finally, (2) taken by itself might be interpreted to permit any risk-imposing activity so long as the actor was willing and able to compensate victims who suffer harms as a result. But imagine someone whose only source of enjoyment comes from playing Russian roulette on people walking by his residence in a crowded city. To make the example more troubling (if not more realistic), imagine that his gun has a million chambers with only one bullet in it and, to avoid the problem of the impossibility of compensating those who might die from his actions, suppose also that he aims only at the pedestrians’ legs. He is very wealthy and willing to compensate anyone who loses a leg as a result of his amusement. Surely, as Sidgwick claims, this is a kind of mischief that we would properly prohibit, and not necessarily because the risk is “too great to run.” We would not feel differently if our imagined agent agreed to add yet another million chambers to his gun.

Our intuitions in such cases may lead us to favor Nozick’s response (1) and say that, just as we properly prohibit activities that cause harm or violate the rights of others unless we have their prior explicit consent (which we will discuss below), so we should prohibit activities that impose the risk of harm or rights violations without the consent of those who bear the risk. But that principle would make life as we know it impossible. When we heat our homes or cook our

meals, most of us cause pollution that increases the risk of harm to others. If we drive a car, we impose risks not only on other drivers but also on pedestrians, some of whom, like children, do not themselves engage in a similar risky activity and so cannot be interpreted as giving their consent to the risk that such activities impose. Moreover, we are not willing to forego altogether the benefits from the mining and manufacturing activities that impose significant risks on many people but produce the goods we consume.

## Permitting Productive Risky Activities

---

Nozick considers another possibility for a principle of compensation for risk. Begin with a person who engages in a productive activity that imposes some risk on others. Someone may start a business that pollutes some nearby water; someone else may drive a car, either to get to work or as part of a business that involves transporting or delivering goods. Because we want to allow such activities, we might argue that they should not be prohibited simply because they impose some risk of harm on others. We might further insist that the person engaged in these activities must prove herself able to compensate others if her risk turns out actually to harm them, perhaps by buying insurance. This proposal is compatible with at least two reasons for prohibiting risk-imposing activities that are congenial to rights-based ethical theories. First, if people oppose being exposed to risks that are imposed as a by-product of productive social activities, either because the probability of harm is too great or because they oppose the nature of the risk, they may decide to prohibit the activity provided they compensate the party who is now prevented from engaging in it (Coase 1960). If people are not willing to compensate actors whose productive but risk-creating activities they wish to prohibit, then they must accept the risk with the understanding that they will be compensated only if they are actually harmed. Referring back to Sidgwick's examples, this proposal would mean that a society that wanted to restrict the manufacture and carriage of explosive substances or prohibit the import of cattle from countries where disease is rife would have to compensate the manufacturers or importers for keeping them from engaging in these activities. The justification for regulating risk in this way, through laws or policies, is that these regulations mimic the behavior of individuals freely engaging in the exchange of goods and services in a market.

The second reason this proposal allows for prohibiting risk addresses nonproductive mischief. The person who wants to play Russian roulette on pedestrians is engaging in a nonproductive activity, which we may prohibit without compensation. In contrast to this example, if we want to prohibit an epileptic from driving, assuming she would impose the same level of risk as the person playing Russian roulette, we would have to compensate her for this loss of liberty. This reason for prohibition without compensation, as Nozick describes it, would also justify prohibiting someone from practicing extortion by proposing to begin a risky activity only in order to obtain compensation from others who want him not to engage in it. Extortion is properly illegal and ethically prohibited. Prohibition relieves the victims from the risk of harm, and the extorter is made no worse off except for being unable to benefit merely from threatening others.

There are nevertheless several problems with this proposal, which sees risk regulation primarily in terms of the justifiable restriction of individual rights and liberties. Describing these problems also highlights the issues that have dominated philosophical discussions of the ethics of risk over the past 40 years.

One problem, already mentioned, is that people cannot be compensated when the harm that results from a risky activity is death. It is not obvious why, on a rights-based theory, people exposed to a risk of death through someone else's productive activity must either accept the risk or compensate the producer to halt it. A second problem with the compensation proposal arises when we consider the distribution of risk and benefit. The proposal insists that risk producers should be compensated for prohibitions that result in their losses. But risk-imposing activities that benefit the creator of the risk may cause harms to others that collectively greatly exceed the benefits to the producer. Why should the victims be required to pay compensation to halt activities that produce more overall harm than benefit? It is not obvious why liberty should have this kind of priority over welfare.

A third problem with this proposal is that it does not help us to deal with "zero-infinity risks," the kinds of activities that produce a very small probability of very great harm, such as nuclear power. When nuclear reactors are properly constructed and maintained, with adequate safeguards, the chance of a catastrophic accident is very low – but it is not eliminated or reduced to zero. And if a catastrophic accident occurs, the producers would be unable fully to compensate everyone who is harmed. According to the proposal we are considering, the failure to be able to compensate is sufficient reason to prohibit the activity. But a society might decide that the benefits of nuclear power – a plentiful supply of reasonably priced and (except for the unsolved problem of permanently disposing nuclear wastes) relatively environmentally benign source of electricity – is a risk worth taking. One of course hopes that few nuclear accidents will occur and that none will be catastrophic, but whether or not nuclear power is a reasonable risk for a society to accept is not something that the rights-based compensation proposal we are considering can help us decide.

Finally, as Nozick himself recognized, the problem of determining what is and what is not a productive activity is not obvious and is inevitably subject to framing effects or how an activity is described. Playing Russian roulette for fun on non-consenting pedestrians is not a productive activity; but engaging in the only kind of activity that one finds enjoyable may be. Restricting our thought to more realistic examples does not diminish this problem. Is allowing people to drive whatever cars they prefer, regardless of the emission of greenhouse gases, a productive activity? Does the reduced regulation of financial institutions, which vastly increased both profits to the banks and risk to everyone else, count as contributing to productive activities? Perhaps "productive" is not the most important concept for determining when societies may reasonably regulate or prohibit certain risk-creating activities.

---

## Risk and Consent

---

Individuals accept risks all the time. We willingly engage in rock climbing, skiing, and other adventures; we put our money at risk for the chance of greater gains; we accept the risk and cost of air travel for the benefit of greater and more convenient mobility; we risk death from surgery for the chance of curing a disease or relieving pain; we get married; we divorce; and so on. Life is full of risks, which we try reasonably to manage far more often than we try to eliminate. Presumably, these risks are acceptable because we consent to them. If we have thought carefully about some prospect, we may reasonably judge that the expected benefits outweigh the cost and risk involved. In some areas, such as medical treatment, the consent process tends to be formal and elaborate. This may be because, in the context of surgery, consent clearly distinguishes a medical procedure from assault with a deadly weapon.

If consent justifies exposure to risk, then the problem for social risk-management decisions and policies, where it is impossible to get unanimous explicit consent, is to find some surrogate, hypothetical, or implicit justification. The practice has been to assume that a policy that monetizes risk, costs, and benefits and can be shown to maximize net benefits is a policy to which rational individuals would consent. But this assumption is not always reasonable, for the measures that show expected net benefits may ignore considerations of justice and the moral demand to show respect for each individual. Relying simply on a social cost-benefit analysis could allow, for example, that the lives or health of some people may be sacrificed for marginal benefits to a great number of other people, because the overall benefits outweigh the overall costs. Instead, the ethical justification for using different methods of risk-cost-benefit analysis to justify a decision or policy depends on showing how the analysis reflects the preferences of those affected by the decision. The justification appeals to the implicit or hypothetical consent of the affected population. This is consent to accept some level of risk in exchange for certain benefits, or an agreement that reducing or eliminating some risk is too costly (MacLean 1986a). If this justification is to succeed, at least two kinds of problems must be solved.

The first problem involves our understanding of individual preferences for risk. Consider a typical case in which we are able to reduce a risk by reducing exposure to some dangerous substance, but the marginal cost increases for each increment of reduced exposure. It may be cost-effective to begin to reduce the amount of mercury in a community's drinking water or to reduce exposure to a carcinogenic substance produced in the manufacturing process at some factory. But the cost typically rises as we continue to reduce these exposures, and it may be prohibitively expensive to try to eliminate them altogether. The question then becomes, "How safe is safe enough?" At what point do people exposed to the risk find it reasonable to accept a given amount of risk rather than pay the cost of further reduction? The easiest way to answer this question would be to monetize the value of an increment of risk, so that we can compare it to other costs and benefits. We explore various methods proposed to do this below, but we must also ask whether all risks – that is, identical probabilities of similar harms – should be valued the same. If we can determine what people generally regard as safe enough in one area, can we generalize this result and apply it to other risks? It turns out that studies show that most people value different kinds or sources of risk differently, so that preferences for risk-cost-benefit trade-offs are to some extent independent of a simple function of probability and magnitude of harm. Individual risk preferences vary among other qualitative dimensions of risk (Slovic et al. 1979). Thus, a person might ride a motorcycle to attend a rally to demonstrate against nuclear power. If he knows what he is doing and is not deemed irrational, then we must conclude that his concern for risk is not merely a function of his perceived probability of a given harm.

So perhaps we should manage risk in a way that recognizes the qualitative differences of their source and nature. But when we learn, for example, that we spend far more to reduce an increment of risk in the workplace than we do to reduce an increment of a similar risk in regulating automobile and highway safety, should we not at least consider reallocating resources to achieve greater or more efficient overall risk reduction? After all, the person who works at the factory may be the same person who drives to and from work each day. The question of what role, if any, qualitative differences in the source of risk should play in risk-management decisions is one that only ethical reasoning and argument can resolve. It remains a central problem in the ethics of risk analysis and risk management.

The second problem in answering the question, “How safe is safe enough?” concerns specifically the risk of death. Recall the suggestion in earlier sections of this chapter that compensation for exposure to risk should be correlated with compensation for the harm involved. Most methods of risk analysis are based instead on finding out what people regard as compensation for being exposed to different levels and kinds of risks. But of course we cannot be compensated for being killed, so how can a compensation test be used to justify decisions that permit imposing a risk of death? This is a problem that governments especially must confront, because their decisions about when it is too costly to reduce some risk often imply that a predictable number of preventable deaths within a large population will be tolerated in exchange for other benefits to society. Thus, in many social risk decisions, the question, “How safe is safe enough?” is equivalent to the question, “What is the economic value of human life?”

Fifty years ago, some economists proposed a solution to this problem, which has shaped the subsequent development of risk analysis (Mishan 1971; Schelling 1968/1984). While no amount of money can compensate a person for being killed, these economists argued that people in fact manage, within some normal range of risk tolerance, to compare the value of small increments of the risk of death to other costs and benefits. If we can discover and use these preferences to make regulatory or policy decisions, then perhaps we can justify risk policies by claiming that they maximize the satisfaction of preferences. In this way, policy decisions are supposed to rest on the implicit consent of those who are affected by them.

Nobody is proposing that we consider the cost of a rescue mission before deciding whether to try to save the life of an identifiable individual who is known to be in danger. That would be to treat human life as exchangeable for other resources in an ethically unacceptable way. But risk policies often cover large populations where the individuals who will in fact be killed or harmed as a result are not identifiable *ex ante*. In some cases, the victims of a risk decision will not be identifiable *ex post* either. When we decide to permit some risk of an accident, for example, and an explosion occurs later that kills some workers, we will know the identities of the victims of that decision, but only after the fact. When we decide whether or how much to reduce an air pollutant that causes a common disease like lung cancer, however, we will never know more than the number of premature cancer deaths saved or lost by our action. We will not know of any victim whether or to what degree the particular pollutant we tolerated caused the cancer. The “value of human life” in these cases is merely a statistical measure aimed at making more reasonable decisions. It is to be applied in contexts where we do not know, at least *ex ante*, who the victims will be. The harm may fall to a few, but the *ex ante* risk may be spread equally to an entire population.

When citizens learn that a proposed regulation will cost a specified amount of money but will save ten lives per year in a population of 100,000, the only thing any of them may know *ex ante* is that if the regulation is put into effect her annual risk of premature death will be reduced by  $1 \times 10^{-4}$ . If the cost is \$100 per person per year in increased taxes to reduce this risk, then the question can be framed as what people are willing to pay for this kind of risk reduction. The same question can be framed to ask whether the social value of one (statistical) human life is greater or less than \$1 million. The social value of human life is thus determined by individual preferences for risk reduction.

When Thomas Schelling first made this argument in defense of a social value of human life, he suggested two methods for uncovering and measuring these preferences, each of which has been developed over the succeeding decades. The *revealed preference method* claims that

individual economic behavior reveals risk preferences. People make decisions about how much to spend on safety equipment for their homes or automobiles; they spend more money for comparable houses in less polluted or less dangerous neighborhoods; they demand more in wages for comparable jobs in more risky environments; and so on (Viscusi 1992). Of course these data may be very rough and approximate, but proponents of revealed preference studies claim that they give us information that can be used to help government agencies make policies that satisfy citizen preferences.

But even making allowance for the roughness of the data revealed by these studies, critics argue that they rely on further assumptions that are rarely warranted. They assume, for example, that people are generally aware of the levels of risk they are facing; they assume that people have genuine options for what they purchase or where they work so that their choices do in fact reveal their preferences; and they assume that consumers, homeowners, and workers are satisfied with their current levels of risk. In addition to these practical difficulties, moreover, philosophically minded critics have also challenged the assumption that consumer preferences accurately reflect our deepest values (Hausman and McPherson 2009; Marglin 1963; Sagoff 2004). These critics claim that we express different things when we act privately as consumers and when we act publicly as citizens. A reasonable person might shop for the least expensive car while simultaneously supporting legislation to make cars safer and more fuel efficient, even if these regulations drive up the price of automobiles. Laws and social policies can express values in ways that are not available to citizens acting privately in the market. Skeptics about this argument might claim that it is easy and not particularly revealing of values to express a preference in a public way when one is not being asked to pay the cost directly, but serious questions remain about the accuracy of reading off a person's values from her market transactions.

An alternative to looking for economic data to reveal preferences is to ask people directly how much they value risk reduction and what they are willing to pay for it. These techniques, which have been called both the *expressed preference method* and the *contingent valuation method*, have the advantage of getting people to focus explicitly on the issue at hand. They have the disadvantage of taking on the difficulties of discovering preferences through surveys and interviews, rather than examining behavior, especially when no real money is at stake. And in determining preferences for risk, contingent valuation methods have the added difficulty of being susceptible to framing effects, which are numerous and often very subtle (Kahneman, et al. 1999). For example, economic theory tells us that the value of a commodity is what a person is willing to pay for it. This means that what a rational agent is willing to pay to obtain a commodity or benefit should be roughly equal to what he would accept in exchange for it, which implies that what a rational person is willing to pay to reduce some increment of risk should be roughly the same as what he would demand in exchange for accepting an increased identical increment of the same risk, within the normal range of risk that we tend to accept and live with. But willingness-to-pay measures for risk reduction tend to differ from willingness-to-accept measures for an increase in risk, often by a very large amount (Kelman 1981). Since most contingent valuation studies are framed in terms of what people are willing to pay to reduce risk, the results of these studies may be seriously biased. The framing problems are difficult, and no widely accepted solution to them has yet been proposed.

There is one further important objection to relying on individual preferences to determine an acceptable social value of human life. Many philosophers would argue that ethically justifiable decisions about tolerating risk-imposing activities that kill people could be made

only through deliberative processes that appeal to ethical reasons and arguments (Broome 2008; MacLean 2009). The techniques of risk analysis that rely on methods for revealing or expressing individual preferences tend to substitute observation and measurement for deliberation and argument. Their defenders claim that these methods are more scientific and thus ethically neutral. Individuals express their preferences, and the risk analyst's job is to measure and aggregate them. But these claims to neutrality can be challenged. It should be clear from the discussion so far that preferences for risk depend on what people accept as full compensation for harm. And, as we have stated, there is no amount of money or other benefits can fully compensate for the harm of being killed. If we know after the fact the identity of the victims of a policy that tolerates some level of risk, then we also know *ex ante* that the victims, whoever they turn out to be, cannot be fully compensated. To ask people what they would demand in compensation for a policy that will result in the deaths of some of them only before the results are known is to demand that they express their preferences in ignorance of the most relevant facts. When all the facts are known, however, the compensation test will fail. So the preference-based argument depends on imposing a degree of ignorance.

This argument further supports the conclusion that the only justifiable way to make decisions that impose a risk of death on a population is through explicit moral deliberation and discussion that results in some agreement that can be interpreted as explicit social consent to accepting a risk. But this kind of deliberation is difficult, and it may not result in agreement. It is perhaps for this reason that proponents of risk analysis are drawn to eliciting or measuring existing preferences for risk reduction and relying on a compensation test under an imposed veil of ignorance.

## Justice and the Distribution of Risk

---

Social risk decisions, like most other social decisions, will not always result in an equal distribution of risks, costs, and benefits. Thus, we confront questions about fairness. If we locate a hazardous waste site in a disadvantaged community (where a “community” can be a neighborhood or a nation), because those who live in the community seek the benefits, say, of added jobs that come with constructing and maintaining the site, or because, being disadvantaged, they demand less in compensation for imposed risk than residents of wealthier communities, the decision may be economically efficient but ethically unfair. Similarly, decisions to engage in activities that benefit us today but impose serious risks on future generations may be popular but unjust. In these and other ways, risk decisions raise familiar issues of distributive justice. We will not discuss these familiar or general issues here.

But risk decisions also can have unique distributive characteristics that are ethically relevant. These can be seen when we distinguish the individual's perspective on a risky prospect from the social perspective. Imagine a society consisting of ten people who are considering a proposal that would increase the risk of death to each person by .10. This, of course, is an unusually great increment of risk, well above what Schelling had in mind in characterizing a normal range. The example is unrealistically stylized in order to highlight for illustration a point about distribution. A risk analyst might try to decide whether the proposal should be adopted by determining what would compensate the average individual for the amount of increased risk. But there are different sources of risk, with different social profiles, that are consistent with imposing a .10 risk on each individual. Consider just two possibilities, which we can call *strongly dependent* and *strongly*

*independent.* In the strongly dependent possibility, one person will be chosen to die, and the risk of being chosen is distributed to each person equally, for example, by drawing straws. We call this a dependent risk because what happens to each individual partially determines what happens to all the others. If we determine that the first person will survive, then the risk to the remaining nine increases; or if the third person dies, the others will survive. Alternatively, a strongly independent possibility may involve imposing a separate .10 risk of death on each person, for example, by having each person draw a ball from an urn that contains one black ball and nine white balls, then replacing the ball and passing the urn to the next person. In this situation, the outcome for each person has no effect on the outcome for any of the others.

Clearly, both of these situations impose an identical and equal risk on each individual, but from the social perspective the risks are different. In the strongly dependent case, there is no social risk – one person in the group of ten will die for sure. In the strongly independent case, there is an *ex ante* expectation that one person will die, but this means that one death is merely statistically more likely than other possibilities. There is some chance (approximately .35) that all ten will survive, and also a small risk (one in  $10^{10}$ ) that everyone in the society will be killed. These different social profiles are clearly relevant to determining what is best from an ethical or normative perspective. Of course, we would need to know many details about the social situation in order to reach a reasonable ethical conclusion. If it is important that the society continues to exist, then the dependent risk proposal might be better. The person who dies might be honored as a martyr. If there are strong bonds of solidarity among all the members of the society, however, then the independent situation might be better. The citizens of this society might believe that it is important to maximize the chance that none of them will die, even if this makes some worse social outcomes possible. The point is that risk can be distributed in ethically relevant ways that may not be apparent and may not be taken into account by techniques that measure only individual willingness to pay and then try to aggregate these results to reach a social decision (Broome 1982; Keeney 1980; MacLean 1986b).

## Current and Further Research

---

Ethical issues involving risk, as we have explained, are different in important ways from issues involving harm and raise significant challenges for traditional moral theories. Moral philosophers started calling attention to these issues in a serious way beginning only in the 1970s, but the ethics of risk has not yet become a well-studied subfield of moral philosophy or a subject that is included in the curricula of most moral philosophy classes. There is reason to think that this situation is beginning to change.

Beginning in the 1960s, mainly in response to concerns about risks inherent in technologies like nuclear power, engineers and decision theorists developed analytical methods of risk analysis to provide information and help in guiding the decisions of regulatory agencies charged risk management. These methods were developed as variations of decisions theory and cost-benefit analysis. Philosophers who followed these developments were primarily those with interests in decision theory or in regulatory and tort law. With few exceptions, the ethical issues were discussed only in the context of environmental policies, where the values and risks are particularly hard to quantify and monetize, (Shrader-Frechette 1980; 1990) or in the context of the philosophy of law, where issues surrounding the use of risk analysis to settle matters of liability received much discussion (Cranor 1997; Perry 2001; Sunstein 2002).

But concerns about technological risk are again beginning to capture the attention of moral philosophers more generally (Roeser and Asvelt 2009). The renewed interest is due in part to developments in environmental ethics and the philosophy of law, but it is also related to the dramatic growth of public awareness of the anthropogenic causes of climate change and the need to take measures now to mitigate risks to future generations. This in turn leads to looking more closely at energy technologies like offshore drilling for oil and carbon sequestration for coal, which exacerbate the risks of climate change for current and future generations, and technologies like nuclear power, which holds out promise of an energy source that does not contribute to climate change but poses other well-known risks. How will research in the ethics of risk respond to these issues? Some of the themes are beginning to come into view. We will end this chapter by briefly describing three of them.

First, debates already underway about the adequacy of formal risk analysis or cost-benefit analysis to guide us to reasonable decisions will continue and intensify. Some philosophers have argued that there simply is no alternative to relying on these formal techniques as a framework for thinking rationally about the ethics of risk (Sunstein 2005). Other philosophers, however, continue to highlight the problems involved in quantifying and comparing all the relevant values that go into an ethically justifiable decision (Hansson 1993, 2003, 2007a, b). Some of these problems involve trying to quantify or apply probabilities to scenarios that are possible but whose likelihood is genuinely uncertain. Philosophers and others are becoming increasingly aware of the need to distinguish risk or frequency from genuine uncertainty.

A second set of issues, especially prevalent in thinking about climate change, arises in considering risks we are imposing on future generations. The ethical problems of dealing with future generations has been a philosophical topic for some time, but the importance of assigning discount rates to risks, costs, and benefits as they occur further into the future is especially dramatic in proposals for responding to climate change (Parfit 1984; MacLean 1983; Broome 2011). The fundamental ethical issue is whether there is any justification for incorporating “pure time preference” into analyses of risks that have long time horizons. The implication of pure time preference is that costs and benefits – including the value of human life – count for less as they occur further into the future. Many philosophers and increasingly some economists argue that pure time preference has no ethical justification. But other philosophers and economists claim that discounting for time reflects the preferences of current citizens, and if our analyses did not include discounting for time, the burden of costs that ought to be borne by the current generation becomes excessively great. Issues about assigning discount rates, which have traditionally been confined to technical discussions of dynamic economic models, are likely to become increasingly central to many ethical discussions.

Third, the uncertainties of many risk issues have led to proposals to constrain risk-benefit analysis as a framework for making trade-off decisions with a more conservative precautionary principle. Again, climate change has been the issue that prompts much of this discussion. In an early effort to coordinate an international response to climate change, the United Nations, in the *Rio Declaration* of 1992, approved a version of a precautionary principle. The *Rio Declaration* states, “In order to protect the environment, the precautionary approach shall be widely applied by States according to their capabilities. Where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation” (UNEP 1992). The idea of a precautionary principle as an ethically justifiable principle to guide policies and laws is the subject of much debate. This principle seems to be more widely accepted on ethical grounds in

Europe and the developing nations than in the United States, where it is widely criticized by philosophers and decision theorists. The heart of the debate seems to be, on the one hand, a distrust of the more formal techniques of risk analysis for dealing reasonably with genuine uncertainties and potential catastrophes and, on the other hand, skepticism on the part of its critics that a precautionary principle can be formulated in a way that can be applied objectively without implying unreasonable conclusions about how we should respond to what might be very remote and unlikely possibilities that someone has imagined, nobody believes is very likely, but which cannot entirely be ruled out.

## Conclusion

---

In this chapter, we have been describing some of the ethical issues that are specific to risk. We have not defended any particular ethical principles, but the discussion does support an important conclusion. Risk analysis has come to be closely identified with the process of estimating risk, determining what people are willing to pay for reducing risk, and using this information to guide government or social policies and regulations. If the arguments above are correct, then the aims of risk analysis should be modest. It can give us important information, but in most cases it cannot be used alone to determine what is an ethically justifiable risk decision. Measuring and aggregating individual preferences for risk is not always an ethically neutral activity, even when it is carefully done. If we are to make decisions that are ethically justifiable, we cannot simply replace ethical deliberation, reasoning, and argument, by techniques that measure and aggregate individual preferences. Those techniques also need to be ethically justified, and considerable work remains to be done to determine the nature of justification in situations involving risk.

## References

---

- Broome J (1982) Equity in risk bearing. *Oper Res* 30:412–414
- Broome J (2008) The ethics of climate change. *Scientific American*, June 2008, pp 97–102
- Broome J (2011) The morality of climate change. WW Norton, New York
- Coase RM (1960) The problem of social cost. *J Law Econ* 3:1–44
- Cranor C (1997) A philosophy of risk assessment and the law: a case study of the role of philosophy in public policy. *Philos Stud* 85:135–162
- Gillroy J (2000) Justice and nature: Kantian philosophy, environmental policy, and the law. Georgetown University Press, Washington, DC
- Hansson SO (1993) The false promises of risk analysis. *Ratio* 6:16–26
- Hansson SO (2003) Ethical criteria for risk acceptance. *Erkenntnis* 59:291–309
- Hansson SO (2007a) Philosophical problems in cost-benefit analysis. *Econ Philos* 23:163–183
- Hansson SO (2007b) Risk and ethics: three approaches. In: Lewens T (ed) *Risk: philosophical perspectives*. Routledge, London
- Hausman D, McPherson M (2009) Preference satisfaction and welfare economics. *Econ Philos* 25:1–25
- Hobbes T (1660/1996) *Leviathan*. Cambridge University Press, Cambridge
- Kahneman D, Ritov I, Schkade D (1999) Economic preferences or attitude expressions? An analysis of dollar responses to public issues. *J Risk Uncertainty* 19:203–235
- Keeney R (1980) Equity and public risk. *Oper Res* 28:527–534
- Kelman S (1981) Cost-benefit analysis: an ethical critique. *Regulation* 5:33–40
- MacLean D (1983) A moral requirement for energy policies. In: MacLean D, Brown P (eds) *Energy and the future*. Rowman & Allanheld, Totowa, pp 17–30
- MacLean D (1986a) Risk and consent: philosophical issues for centralized decisions. In: MacLean D (ed)

- Values at risk. Rowman & Allanheld, Totowa, pp 17–30
- MacLean D (1986b) Social values and the distribution of risk. In: MacLean D (ed) Values at risk. Rowman & Allanheld, Totowa, pp 75–93
- MacLean D (2009) Ethics, reasons and risk analysis. In: Roeser S, Asveld L (eds) The ethics of technological risk. Earthscan, London, pp 115–127
- Marglin S (1963) The social rate of discount and the optimal rate of investment. *Q J Econ* 77:95–111
- Mill JS (1859/1978) On liberty. Hackett, Indianapolis
- Mishan EJ (1971) Evaluation of life and limb: a theoretical approach. *J Polit Econ* 79:687–705
- Nozick R (1974) Anarchy, state, and Utopia. Basic Books, New York
- Parfit D (1984) Reasons and persons. Oxford University Press, Oxford
- Perry S (2001) Risk, harm, and responsibility. In: Owens D (ed) Philosophical foundations of tort law. Oxford University Press, Oxford
- Roeser S, Asveld L (eds) (2009) The ethics of technological risk. Earthscan, London/Sterling
- Sagoff M (2004) Price, principle, and the environment. Cambridge University Press, New York
- Schelling T (1968/1984) The life you save may be your own. In: Chase S (ed) Problems in public expenditure analysis, The Brookings Institution, Washington, DC; reprinted in Schelling T, Choice and consequence: Perspectives of an errant economist, Harvard University Press, Cambridge MA
- Shrader-Frechette K (1980) Nuclear power and public policy: the social and ethical problems of fissile technology. D. Reidel, Dordrecht
- Shrader-Frechette K (1990) Ethics and risk-benefit analysis. Ethical and policy issues perspectives 9:6–7
- Sidgwick H (1904) The elements of politics, ch. 9, sec. 2. Macmillan, London
- Slovic P, Fischhoff B, Lichtenstein S (1979) Rating the risks. *Environment* 21(14–20):36–39
- Sunstein C (2002) Risk and reason. Cambridge University Press, Cambridge
- Sunstein C (2005) Laws of fear: beyond the precautionary principle. Cambridge University Press, New York
- UNEP (1992) Report of the United Nations conference on environment and development. <http://www.un.org/documents/ga/conf151/acnfl5126-1annex1.htm>. Viewed 12 June 2011
- Viscusi WK (1992) Fatal tradeoffs: public and private responsibilities for risk. Oxford University Press, New York

# **31 Toward a Premarket Approach to Risk Assessment to Protect Children**

*Carl F. Cranor*

University of California, Riverside, CA, USA

<i>Introduction</i> .....	<b>806</b>
<i>History</i> .....	<b>806</b>
<i>Risk Assessment and Postmarket Laws</i> .....	<b>807</b>
<i>The Need for Premarket Toxicity Testing and Risk Assessment</i> .....	<b>811</b>
<i>Further Research</i> .....	<b>815</b>
<i>Conclusion</i> .....	<b>816</b>

**Abstract:** Postmarket laws and risk assessments seemed to be a good idea in the 1960s and 1970s when the public and Congress in the United States were widely alerted to environmental health harms. And, they were an improvement on existing regulatory laws, of which there were few, and an improvement over torts or the criminal law to protect the public's health.

By now, scientific developments have revealed the substantial shortcomings of such laws. Given the number and kinds of diseases and dysfunctions that may be attributable to in utero or early childhood contamination, postmarket laws and risk assessments will no longer serve to protect children. Premarket toxicity testing and any needed risk assessment before commercialization will have to supplant existing postmarket laws and risk assessment.

## Introduction

---

Historically in the United States, and very likely in many other countries, risk assessment has been combined with postmarket legislation which together aim to protect the public's health from hazardous substances. Postmarket laws permit 80–90% of industrial chemicals to enter commerce *without any legally required testing*. Risk assessment is “the use of the factual base to define the health effects of exposure of individuals or populations to hazardous materials and situations” (NRC 1983).

Because of recent scientific developments in understanding disease processes showing that some diseases and dysfunctions begin in the womb or in childhood, postmarket laws combined with risk assessments fall quite short of protecting the public from toxicants. Consequently, the use of risk assessment and its legal context needs to be reconceived. To the extent that risk assessment has a significant role to play in public health protections in the future, I argue that it should be much more a part of premarket assessment of the toxicity of products in legal contexts in which substances are tested for their toxicity before entering commerce. However, postmarket risk assessments likely will not disappear, because there will continue to be some use for them in limited circumstances, even after laws are amended to require adequate premarket toxicity testing for industrial chemicals.

## History

---

In the late 1960s and early 1970s when environmental concerns came to the forefront of public awareness, the U.S. Congress (and legislatures in many states) sought to address environmental and environmental health issues. These legislative bodies passed various kinds of legislation aimed at protecting people's health and the environment. At the time, this legislation endeavored to remedy shortcomings of existing institutions and to address better and more comprehensively risks to public health and the environment.

The US Congress passed the Occupational Safety and Health Act aiming to protect employees from workplace harms and toxicants. It also sought to protect against pollutants posing risks in the air, surface waters, or drinking water. The Consumer Product Safety Act addressed risks from consumer products. It also amended the pesticide laws and the laws concerning food additives, food contaminants, and drugs (U.S. Congress, OTA 1987).

By 1976, legislators recognized that these efforts were inadequate because they tended to focus on pollutants one media at a time – surface waters, or drinking water, or the air, or the workplace – or they addressed a small number of substances with each law, for example,

pesticides, drugs, food additives, and consumer products. A major shortcoming was that media-by-media or venue-by-venue legislation did not address toxicants more comprehensively. Thus, in 1976 Congress sought to close this gap with the Toxic Substances Control Act (1976) (Applegate 2009). It also passed various laws to guide safer creation, use and disposal of toxicants from cradle to grave, and to clean up toxicants that had been poorly disposed in the ground or groundwater (U.S. Congress, OTA 1987).

As a result of this legislation, “postmarket” laws govern a very large percentage of substances in the market. Thus, Congress authorized public health agencies to “police” exposures to substances once they are in the market to try to identify substances that pose risks to the public or workforce (Merrill 2004). Similar to a policeperson identifying a legal violation and issuing a citation or arrest warrant for the offender, a regulatory agency has a legal burden of proof to identify a risk, develop a scientific case that exposures to the substances pose risks, and then produce a legally binding regulatory response to reduce or eliminate the risks before (any or too much) harm occurs (Wagner 2008). Traditional risk assessment is the key to postmarket prevention of harm (more on this next section).

Other products are subject to premarket testing and approval laws. For pharmaceuticals, new food additives, and pesticides – a relatively small number or substances (about 10%, perhaps up to 20% of the chemical universe) – Congress required companies seeking to commercialize such products to test them for various adverse effects *before* they can enter the market. Once those tests are completed the company must petition an agency, such the Food and Drug Administration (for drugs and new food additives) or the U.S. Environmental Protection Agency (for pesticides) to review the test data and to license the company to sell its product in the market. These are “pre-market testing and approval laws” (U.S. Congress, OTA 1987).

## Risk Assessment and Postmarket Laws

---

How can postmarket laws “prevent” harms? That is, they do not authorize legal action against substances that may be toxic until after they are in commerce and pose risks to the public. How can these be protective?

The critical idea is the concept of a risk. Risks are merely the *chance* of harm or some other untoward or undesirable outcome occurring (Rescher 1983; NRC 1983). By definition a risk has not yet resulted in harm, but is merely a possible or probable adverse outcome. Consequently, if there were reliable techniques for identifying risks before they materialize into harms, Congress could authorize agencies to be alert for products that pose *risks*, identify those as early and as quickly as possible, and then expeditiously reduce or eliminate them before they caused harms. The agencies typically would be authorized to “issue a rule” that provides the health protections against risks required by law in question.

Some laws might require that an agency identify substances that posed risks to human health and then identify technologies that would reduce the risks as much as the technology permitted. These are so-called technology-based laws. Other laws might require an agency to identify risks to human health and then issue public health protections that reduced exposures in the environment to a legally mandated level, for example, so that it did not constitute an “unreasonable risk,” (typical of pesticide laws or the Toxic Substances Control Act) or so that it provided health protections to the public with “an ample margin of safety,” taking vulnerable subpopulations into account (the Clean Air Act) (U.S. Congress, OTA 1987).

The focus on risks was typically supported by experimental animal studies and other nonhuman scientific data so humans would not need to suffer harm in order for an agency to develop health protections (U.S. Congress, OTA 1987). In principle studies conducted on animals would reveal potential risks to people so exposures could be reduced before harm occurred, although the use of animal data has become sufficiently controversial in some areas, for example, in Europe in particular, that governments are trying to find alternatives to animal studies for this purpose.

Congress also seemed to envision that a public agency could issue a health regulation to protect citizens from a toxicant fairly quickly. Overall, this approach seemed to be a clever idea. As a result of legislation targeting risks identified by animal and other nonhuman studies, a new field of risk assessment was born. If the laws and their procedures functioned well, the identification of risks followed by quick regulatory action could prevent harm. In fact, early on some risk assessments were done comparatively quickly (U.S. Congress, OTA 1987).

Some legislation imposes much greater legal burdens on public health agencies than other laws. *Ambient exposure laws* require public health agencies to determine whether existing exposures cause risks to the public, and then assess what levels of exposures, given the potency of the substance, will result in an acceptable risk under the legislation. Conducting a risk assessment under ambient exposure laws requires all four stages of a full-fledged risk assessment. Public health agencies first must identify substances that are hazardous. Hazard identification consists of the “determination of whether a particular chemical is or is not causally linked to particular health effects” (NRC 1983). The idea is whether the chemical has intrinsic “built-in ability[ies] to cause an adverse effect” (Faustman and Omenn 2001; Heinzw 2009). Next, they must determine what exposure levels can cause adverse effects. This is the estimation “of the relation between the magnitude of exposure and the probability of occurrence of the health effects in question” (NRC 1983). This is sometimes called a “potency assessment,” because it reveals the toxic potency of the substance.

Third, health agencies would conduct an exposure assessment. This is the “determination of the extent of human exposure before or after application of regulatory controls” (NRC 1983). Typically, this would constitute the level of exposures in the relevant environment at which people would be exposed. Finally, agencies must provide an overall risk characterization. This is the “description of the nature and often the magnitude of human risk, including attendant uncertainty” (NRC 1983; Cranor 1993).

In contrast to ambient exposure laws, when the U.S. Environmental Protection Agency must protect the public health under a *technology-based statute*, such as the Clean Air Act or Clean Water Act, its risk assessment steps are comparatively minimal. The EPA need only conduct a *hazard* assessment, that is, to determine whether the chemical in question has intrinsic toxicity properties that can cause adverse health effects. It need not go through the other three risk assessment steps. Once it has found that a substance has such toxic properties, such as being a carcinogen or a reproductive toxicant, it would typically require technology to bring exposures to the chemical to the lowest level the technology can achieve (U.S. Congress, OTA 1987). The moral view underlying such legislation is that companies must “do the best they can” to reduce exposures to toxicants with existing or achievable or the best technology (different laws have different standards) (Wagner 2008).

Risk assessments under *ambient exposure laws* are much more data intensive, science intensive, labor intensive, and time consuming. A public health agency must conduct all four steps in typical risk assessments, and it might even have to provide two risk characterizations:

One before exposures are reduced and then one assuming reduction in exposures with the second used to show that steps taken to protect the public health would indeed do so and how much protection would be achieved.

Unfortunately, postmarket risk assessments to protect the public health have not lived up to the proposed ideal. Postmarket regulatory processes are not quick. Agencies do not quickly find or develop data from animal or other studies. There are numerous incentives in postmarket laws and procedures for affected companies to slow the process. Agencies also have limited personnel to devote to developing protections because they are too underfunded. Consequently, these efforts have bogged down in scientific and other disputes. Issuing health standards to better protect the public can easily take several years, sometimes much longer and occasionally decades, as it has so far with regard to dioxin in the US. Legislation based on postmarket regulation is especially inadequate for protecting our children (Cranor 1993).

When there is insufficient scientific data about toxic hazards from products under postmarket laws, public health agencies find it difficult to provide better health protections. They need enough data to identify risks before technology-based laws can be issued. And these agencies must have even more evidence to issue protective ambient exposure standards. These circumstances create temptations for companies not to test their products for toxicity. If there is no data, a company could say that it had no evidence of adverse effects. If the EPA seeks to require a company to provide the data, it must go through a burdensome legal process to do so (Cranor 2011).

A second postmarket temptation, utilized as early as the 1970s by the tobacco industry, is to raise doubt about the science that shows the toxicity of a product (Michaels 2008). Since a public health agency has a legal burden of proof to find and assemble data about the toxicity of a substance, a company has strong incentives to challenge the data. As an internal memo from a tobacco company expressed it, “Doubt is our product since it is the best means of competing with the ‘body of fact’ that exists in the minds of the general public. It is also the means of establishing a controversy” (Brown and Williamson Tobacco Company 1969). Casting doubt on the science also challenges any attempt to improve health protections, because it gives the appearance that the science is not settled, even if it largely has been. And this tactic can substantially delay better health standards. Moreover, challenging scientific results can be done without disagreeing with public health standards, an important public relations point. The end result is a significant delay in health standards (Cranor 2011).

In addition, companies can be tempted to demand “proof” of toxicity from public health agencies before exposures are reduced (Michaels 2008; Cranor 2008). The idea of proof is more typical of deductive math, where logical certainty is required, but is an alien idea in the biological sciences, where public health officials typically make decisions based on the weight of the scientific evidence. To the extent a party subject to regulation can persuade a regulatory agency that the agency must meet very high standards of proof before it modifies the legal status quo, this can erect obstacles to improved health protections even higher than the particular statute in question may require.

Companies might also urge public health agencies to require myriad kinds of evidence in support of better public health protections, for example, mechanistic data about how a substance functions at the molecular level. This is usually evidence that is very difficult to provide and ordinarily that is simply not known. For example, scientists understood numerous adverse and beneficial effects of aspirin for approximately 100 years without understanding the mechanisms by which either result was achieved (Santone and Powis 1991, p. 169). Or if there

were some epidemiological evidence showing associations between exposures to toxics and adverse effects, companies might still argue that there should be more epidemiological studies to corroborate existing data. Even if there were human data showing adverse effects, firms might argue that there should be animal studies as well in order to have a model for toxicity effects.

Such requests might be appropriate in the world of academic science, where issues of modest moment could be at stake. However, when protection of the public's health is at stake unreasonable demands for scientific studies can have undesirable consequences. When a product is hazardous and poses risks, such additional demands leave the public at risk while studies are carried out because if substances pose risks or are actually harmful this risks type II errors (not treating a toxic substance as toxic) [See Hansson on type I and type II errors]. Of course, an agency must have appropriate data to support its recommendations for improved health standards. However, excessive demands for more evidence, spurious claims criticizing the science, and demands that conclusions must be supported with greater certainty all delay health protections. Postmarket laws create all these temptations (Cranor 2011).

Regulated industries have also sometimes utilized less creditable means for opposing improved public health protections. Company experts sometimes mislead public health agencies and scientific journals about the science. Scientists or their employers have sometimes modified studies so that they can claim that there are no or only minimal risks from their products (Michaels 2008).

As David Michaels has noted, companies may "salt the [scientific] literature with questionable reports and studies... which regulatory agencies have to take seriously." However, he argues that the major goal of such studies is "to clog the [regulatory] machinery and slow down the process" (Michaels 2008). At other times, research projects have been proposed and seemingly designed to find no adverse health consequences, even if independent scientists have found they exist. For instance, the petroleum industry was quite explicit that studies assessing whether benzene contributed to blood and bone marrow cancers would be designed to find results that could be used to fight health protections for its workers and used in fighting toxic tort suits (Cappiello 2005; Bohme et al. 2005).

Finally, postmarket laws have resulted in a chemical world in which little is understood about the toxicity of industrial chemicals to which we are all exposed. Since the vast majority of substances are subject to postmarket regulation, there is no toxicity data for about 70% of these substances. In 1984, about 75% of 3,000 substances produced in the highest volume and likely having the greatest exposures had no toxicity data (NRC 1984). There has been slight improvement since that time, but only very recently and it appears there is no toxicity data for about 50% of the substances produced in the highest volume.

In addition, for about 50,000 new substances entering commerce since 1979 approximately 85% lacked health effects data. For a majority of 50,000 even their chemical identity is hidden by confidential business information claims (Guth et al. 2007, pp. 237–238). If there were an effort to test substances already in commerce at the rate such tests occur this could take hundreds of years. Moreover, many substances are more toxic than might appear. A sampling of 100 chemicals in commercial use and a sampling of 46 chemicals produced in more than a million pounds revealed that about 20% of each group was mutagenic. Mutagenicity tests are comparatively quick and inexpensive to conduct, yet these have not been performed on most substances. A mutagen has a substantial likelihood of being a carcinogen in mammals (Claxton et al. 2010).

The above discussion suggests that postmarket laws combined with risk assessments utilizing nonhuman data are inadequate to protect human health well. If human data were required as companies sometimes urge, this would defeat the purpose of disease prevention. This concern is even of greater urgency when one understands some new developments in biology called the developmental origins of disease.

## The Need for Premarket Toxicity Testing and Risk Assessment

---

Recent scientific developments have revealed that people are extensively contaminated by a variety of human made chemicals substances, and that because of contamination during development fetuses and children can experience disease, dysfunction, and death as a consequence. Developing children are at even greater risk than adults.

Each of us is contaminated by hundreds of substances. Industrial chemicals, pesticides, cosmetic ingredients, and other products can enter our bodies and then infiltrate our tissues, organs, and blood. The US Centers for Disease Control (CDC) has developed techniques for detecting them by measuring the amounts in our blood or urine; this process is called biomonitoring (U.S. Department of Health and Human Services, CDC 2005).

Biomonitoring directly measures the “levels of chemicals that actually are in people’s bodies” (Sexton et al. 2004). This procedure reveals “which chemicals get into Americans and at what concentrations,” the percentage of people exposed to toxic concentrations, and the range of body burdens (U.S. Department of Health and Human Services, CDC 2005). It can also provide an integrated effect of “all routes of exposure – inhalation, absorption through the skin and ingestion, including hand-to-mouth transfer by children” (U.S. Department of Health and Human Services, CDC 2005).

At present, the CDC can reliably identify about 212 substances in US citizens’ bodies. As protocols for detecting industrial chemicals are developed, this number will certainly increase because all but the very largest molecules will enter our tissues (Needham 2007; U.S. Department of Health and Human Services, CDC 2009). The CDC investigates these compounds because they are known or suspected hazardous substances or represent substantial exposures to US citizens (U.S. Department of Health and Human Services, CDC 2005; Heinzow 2009; Faustman and Omenn 2001). Whether hazardous compounds will pose risks depends upon other considerations such as their concentration in people’s bodies.

The substances CDC is investigating include the gasoline additive, methyl tertiary butyl ether (MTBE), perchlorate, a discarded rocket fuel component, polychlorinated biphenyls (PCBs) that were used as industrial insulating and cooling compounds, polybrominated diphenyl ether fire retardants (PBDEs), numerous pesticides, perfluorinated compounds (PFCs) that were used in Teflon, among other products, as well as arsenic and lead. (U.S. Department of Health and Human Services, Fourth Annual Report, U.S. Department of Health and Human Services, CDC 2009). Most of these compounds can cause toxic responses in animals or humans at some level of exposure: estrogen-mimicking substances and other hormone-disrupting chemicals; thyroid disruptors; neurotoxicants, which can damage the nervous system; developmental toxicants; immunotoxins, which can damage the immune system; as well as known or probable human carcinogens.

Some of these substances are persistent and will remain in people’s bodies for years. PCBs have a half-life of about 8 years. The half-life of a substance is the time it takes for one-half of its

concentration to leave the body. In addition, perfluorinated substances have a half-life of about 7 years (Calafat et al. 2007). The PBDE flame-retardants can reside in fatty tissues for several years. Together PCBs and PBDEs can cause neurological effects including behavioral disorders (Costa and Giordano 2007). These two plus perchlorinated substances affect different neurological pathways that could contribute to similar problems (Woodruff et al. 2008). Because each has a long half-life in human bodies, they have an extended period to adversely affect children's health. For example, experimental animals studies PFCs and PBDEs appear to have neurotoxic effects, causing deranged behavior in rodents (Eriksson et al. 2009).

Some substances including bisphenol A are transient in human bodies. However, studies reveal that more than 90% of the populace older than 6 are contaminated by BPA, with children being the most highly contaminated (Calafat et al. 2008). This finding indicates that US citizens have almost continuous exposures. Other kinds of substances that have short half-lives in people's bodies include the gasoline additive, MTBE, other solvents as well as plastic additives such as phthalates.

Whether persistent or transient, toxic substances can pose heightened risks to children during development. Because children have increased vulnerability during development their risks from toxicants are heightened. In addition, many substances in a pregnant woman's body will cross the placenta and enter her developing child. Many of the same compounds will also contaminate her breast milk and be transmitted to her child through nursing. Whether in utero or through nursing she will offload some of her body burden of industrial chemicals onto her child. Newborns have up to 232 industrial chemicals in their bodies (Fimrite 2009). Thus, industrial compounds sully children prenatally, and postnatally this body burden is increased from external exposures, including mother's milk. Some of the same compounds may be found in formula milk, but formula could also contain other substances of concern.

Perhaps this would be of lesser concern, if developing children were no more susceptible to diseases than adults. However, they tend to be more susceptible to diseases, dysfunctions, or premature death as a result of prenatal or early postnatal exposures. They are typically more highly exposed than adults both prenatally and immediately postnatally, but at the same time have lesser defenses against toxicants. They also have more of their lifetime for diseases to manifest themselves (Cranor 2011).

Thus, researchers have documented a number of developmental toxicants. Thalidomide and anticonvulsive drugs cause morphological defects obvious at birth. Other toxicants contribute to long-delayed, but typically less visible adverse effects. Lead can lessen children's IQ and contribute to aggressive behavior (Lanphear 2005) [In adults and at low levels it can also cause cardiovascular risks and strokes (Navas-Acien et al. 2007; Silbergeld and Weaver 2007)]. In utero exposures to pesticides, probably lead, and other neurological toxicants appear to hasten Parkinson's disease decades after contamination. Estrogen-mimicking substances have been documented as leading to early onset of cancer (from diethylstilbestrol [DES]), and some appear to increase the chances of its occurring, or contribute to other diseases, but years after exposure. BPA crosses the placenta, becomes more toxic in the fetal environment, changes how important genes are expressed, and damages placental cells (Balakrishna et al. 2010, p. 393. e6; Bromer et al. 2010). Experimental studies also suggest that it can contribute to breast, uterine, and prostate cancer as well as obesity and diabetes (Heindel 2008). It can also adversely affect the immune system, probably permanently (Dietert and Piepenbrink 2006).

Some known developmental toxicants (in addition to the above) include mercury, sedatives, arsenic, tobacco smoke, alcohol, and radiation. Two hundred known human neurotoxicants that adversely affect adults are likely toxic to children as well. Based on animal research other industrial

chemicals are of substantial concern, including phthalates, PBDEs, and cosmetic ingredients (Grandjean and Landrigan 2006). Many carcinogens, lead, tobacco smoke, disinfectant by-products, and radiation appear to have no known safe level (Wigle and Lanphear 2005).

These contributors to disease have been revealed comparatively recently, although the developmental toxicants, thalidomide, methylmercury, and DES, have been known for more than 40 years (Cranor 2011). Many of the diseases that once plagued the United States have disappeared as a result of public health advances and vaccinations. Cleaning up water and sewage reduced cholera and typhoid fever. Vaccines reduced or eliminated other diseases, including smallpox, polio, measles, and mumps. Children born today are likely to exceed their ancestor's lifespan by 2 decades (Landrigan 2009).

Consequently, although most people are living longer, more subtle diseases remain caused by exposure to various toxicants. These more subtle adverse effects might not have been seen when people had shorter life spans or died from more obvious diseases.

Nonetheless, there remain various diseases, including “cardiovascular disease, diabetes, obesity, respiratory ailments, and injuries” that are of concern. As the above discussion suggests there are also a number of morbidities that can affect children. Some are “intellectual impairments” such as “lowered IQ, poor memory, mental retardation, autism, attention deficit hyperactivity disorder (ADHD).” There are also documented “behavioral problems, asthma, and preterm birth...” (Lanphear 2005). Both groups, according to Lanphear, appear to be related to low-level exposures “to... noninfectious environmental factors or gene-environment interactions” (Lanphear 2005). In addition, obesity, which might be thought to result from behavior or unhealthy foods, can also have contributions from toxicants that enter our bodies without our choice.

For many of these diseases recent research has focused on what are called the “developmental origins of health and disease.” Since diseases typically result from a combination of genetics and environmental influences, and since a person’s genetic code does not undergo rapid change, researchers believe that environmental influences, including particularly molecular exposures, contribute to these adverse effects. Consequently, they are investigating the view that some contaminants, while not modifying a person’s genome, affect its functioning, or how it “expresses itself.” Such effects on the genetic code are the so-called epigenetic effects. Environmental influences, including early life contamination by toxicants, on how genes are expressed, they find, can lead to disease (Jirtle and Skinner 2007; Heindel 2008).

Existing postmarket laws that permit these conditions are reckless toward the public. The policies in the legislation appear to reflect lack of thought about possible undesirable consequences. If companies that commercialize products without testing for toxicity are aware of the contamination of adults and the vulnerability of developing children, they also act recklessly.

The scientific discoveries of contamination and disease in the context of postmarket laws suggest that if a community seeks to reduce risks to its children, the law must change to require appropriate premarket toxicity testing and review by a public health agency before the substances enter commerce. This suggestion resembles in some respects laws governing pharmaceuticals or pesticides in the United States. Most risk assessments that would be done would also be premarket, before industrial chemicals entered commerce. This approach would be somewhat different from postmarket risk assessment. It is also a much more precautionary approach to preventing diseases (See [Chap. 38, The Precautionary Principle](#), in this handbook).

Consider a few of the issues. If premarket toxicity testing revealed that a substance was hazardous, that is, had intrinsic toxic properties, a next step would be to inquire about its potency and whether it operates by means of a threshold or a nonthreshold mechanism. That is, is there some level of exposure to developing children below which the substance would not be toxic or is there no lowest level below which it is toxic, similar to lead or radiation? Moreover, this question should be pursued assuming that pregnant women are contaminated by other substances, some of which might well plausibly interact with the new industrial chemical. That is, toxicity testing should be pursued, assuming the world as it is, namely, that people are already contaminated by other industrial chemicals. In addition, researchers should allow for wider genetic variation in humans than in animal models or test tube experiments in which products are tested (Hattis and Barlow 1996; Woodruff et al. 2008; NRC 2006). They should assume that people including both pregnant women and developing children are already exposed to other toxicants and that they exhibit the full array of susceptibility of which researchers are aware (Cranor 2011).

In addition, to protect developing children researchers should determine whether the substance can cross the placenta or enter breast milk. If it is intrinsically hazardous, is a nonthreshold toxicant, and can cross the placenta or enter breast milk exposing developing children, this is a strong presumptive reason to prevent its commercialization. Testing under such circumstances should both approximate real-world conditions and better protect people and their children from industrial substances that were toxic and proposed for commercialization.

The next step in a premarket risk assessment would be to determine what exposure levels, if any, could be permitted once a substance were in commerce consistent with protecting developing children. For postmarket risk assessments, exposure assessments would seek to determine what existing exposure levels were. For premarket risk assessment, what exposures could be permitted consistent with protecting developing children? Premarket risk assessments would assess some overall characterization of what risks, if any, would exist under exposures that would provide substantial protections to developing children.

Premarket testing and any needed risk assessment would change incentives so that private companies are more likely to act in the public interest than frustrating it as currently occurs under postmarket laws. There would be no incentives to delay the production of scientific data; quite the contrary, companies would seek to have their products enter commerce as quickly as possible. There would be no insistence on multiple studies to document risks or harm. Instead, companies would seek to reduce the number of studies and data requirements to lower costs. Companies would not likely demand that high standards of proof must be satisfied before products could enter the market. More likely, they would argue for lesser standards of proof.

There would still be temptations to produce false or misleading toxicity data as sometimes occurs with premarket pharmaceutical or pesticide testing. Public health agencies would need to be alert to such problems. Moreover, a testing and approval process would need to be designed to minimize this to the extent that it could. For instance, companies might transfer funds to a public health agency sufficient to support needed testing and the agency would then pick a laboratory with a reputation for using credible science to carry out needed testing (Cranor 2011). The institutional structure needed for such purposes would need to be developed in some detail, which I do not have the expertise to do. However, the National Research Council has sketched some of the main features (but even there more details will be needed) (NRC 2008).

Another aspect of premarket testing should consider pollutants. Typically, since risk assessments are usually conducted after toxic pollutants have affected a geographic area, there is no reason that risk assessments could not be conducted in advance of exposures in

order to assess and reduce risks. To illustrate this consider a recent report about pollutants from the Tacoma copper smelter. The EPA had to conduct a risk assessment to determine *after the population was exposed* what concentrations of toxic substances, such as lead, arsenic, and cadmium, had been deposited on the land and homes and how contaminated the populations were, and what risks these posed (State of Washington 2005). The EPA needed this post-deposition risk assessment to assess and clean up the surrounding community. It sought to determine how much damage was being done to the population and the wider environment and how stringent should the pollution emissions be.

Legislatures, now alerted that such problems can occur, should require risk assessments of potential pollutants from industrial facilities *before* damaging effects occur. Reasonable risk assessments could be undertaken in advance by companies to determine how the population would likely be affected, whether the company would need to reduce effluents, and whether public health protections would need to be undertaken as a consequence.

If legislatures required appropriate premarket toxicity testing for new and existing products as well as for major pollutants from industrial facilities, labor-intensive, science-intensive, and slow risk assessments that at present obstruct public health protections under postmarket laws would be needed much less. In the current legal environment, public health agencies employ the scientific community to study health problems or risks with experimental or epidemiological studies or both, assess exposures, and then digest the data to judge public health risks. Premarket testing of products in principle should reduce such time-consuming postmarket studies.

Scientists would continue to have plenty of studies to conduct, but most of them would need to be performed prior to industrial chemicals entering the market. Researchers would still need to conduct tests on experimental animals in order to test for carcinogens or developmental toxicants, for example (at least until these were superseded by more efficient toxicity tests). They would also need to conduct many short-term tests, as well as developing new ones both to avoid animal studies and to find better and quicker means to alert companies and public health agencies to toxicants. There would even be a continued need for exposure assessments, only they would more closely resemble preexposure pollution assessments (described above), during which researchers would need to estimate as best they could what exposures were likely to occur so that they could provide, if needed, an overall estimate of risks from substances if they were commercialized.

It seems to me that there is also a profound paradigm shift that must accompany the enterprise of science in a world in which much greater premarket toxicity testing occurs. Scientists that currently conduct postmarket risk assessments may be committed to studying an existing toxicity problem, discovering its scope, and determining the degree of toxic potency in order to protect the public's health. This paradigm will seem appropriate if most of the risk assessments are conducted before products enter the market. Researchers likely will need to rethink somewhat how they perform studies for premarket risk assessments (but there would be considerable similarities between premarket and postmarket studies) and their sources of funding might be different.

## Further Research

Researchers should be asking what tests and procedures should be used to identify toxicants before they enter commerce and so that diseases do not result from contamination in utero or

perinatally. The emphasis should be on avoiding such risks in the first place. If premarket testing and risk assessments were successful, there would be fewer existing problems to identify and solve, but there would be toxicants to identify before they become public health problems that needed identification and solutions.

Answering these questions will require additional research. Some of it is suggested in the immediately proceeding paragraphs. The generic idea would be to conduct testing and any needed premarket risk assessment to identify risks to developing children before substances enter commerce. For example, since currently we are all contaminated by various industrial chemicals, some toxic, testing should incorporate current contamination during animal studies when a new substance is considered for introduction into commerce. Will the new substance interact with existing contaminants to increase risks, particularly for children? For example, researchers from Andreas Kortenkamp's lab have found this for synthetic estrogens as well as some anti-androgens (Hass et al. 2007; Kortenkamp 2007). Moreover, Woodruff et al. have reported that even when substances do not act by particular cellular receptors, they can affect the same endpoints, for example, reduce thyroid production, which can damage the neurological development of children in utero (Woodruff et al. 2008). Researchers designing testing protocols for premarket testing and any needed risk assessment will need to be imaginative in creating tests that mimic actual exposures to developing children and any other contamination that they might have because of their parents existing contamination.

## Conclusion

---

Postmarket laws and risk assessments seemed to be a good idea in the 1960s and 1970s when the public and the U.S. Congress were widely alerted to environmental health harms. And, they were an improvement on existing regulatory laws, of which there were few, and an improvement over torts or the criminal law to protect the public's health. (Moreover, to the extent that other countries developed similar laws, they too will have the same problems that are now apparent in the United States.)

By now, scientific developments have revealed the substantial shortcomings of such laws. Given the number and kinds of diseases and dysfunctions that may be attributable to in utero or early childhood contamination, postmarket laws and risk assessments will no longer serve to protect children. Premarket toxicity testing and any needed risk assessment before commercialization will have to supplant existing postmarket laws and risk assessment.

## References

---

- Applegate JS (2009) Synthesizing TSCA and REACH: practical principles for chemical regulation reform. *Ecol Law Q* 35:101–149
- Balakrishna B, Henare J, Thorstensen EB, Ponnampalam AP, Mitchell MD (2010) Transfer of bisphenol A across the human placenta. *Am J Obstet Gynecol* 202:393.e1–393.e7
- Bohme SR, Zorabedian J, Egilman DS (2005) Maximizing profit and endangering health: corporate strategies to avoid litigation and regulation. *Int J Occup Environ Health* 11:338–348
- Bromer JG, Zhou Y, Taylor MB, Doherty L, Taylor HS (2010) Bisphenol-A exposure in utero leads to epigenetic alterations in the developmental programming of uterine estrogen response. *FASEB J* 24:1–8, located at [www.fasebj.org](http://www.fasebj.org). Accessed 20 Feb 2010
- Brown and Williamson Tobacco Company (1969) Smoking and health proposal Brown and

- Williamson document no. 680561778–1786, [legacy.library.ucsf.edu/tid/nvs40f00](http://legacy.library.ucsf.edu/tid/nvs40f00). Accessed 30 June 2008
- Calafat AM, Wong L-Y, Kuklenyik Z, Reidy JA, Needham LL (2007) Polyfluoroalkyl chemicals in the U.S. population: data from the national health and nutrition examination survey (NHANES) 2003–2004 and comparisons with NHANES 1999–2000. *Environ Health Perspect* 115:1596–1602
- Calafat AM, Ye X, Wong L-Y, Reidy JA, Needham LL (2008) Exposure of the U.S. population to bisphenol A and 4-tertiary-octylphenol: 2003–2004. *Environ Health Perspect* 116:39–44
- Cappiello D (2005) Oil industry funding study to contradict cancer claims. *Houston Chronicle* April 29:A1
- Claxton LD, Umbuzeiro GA, DeMarini DM (2010) The *salmonella* mutagenicity assay: the stethoscope of genetic toxicology for the 21st century. *Environ Health Perspect* 118(11):1515–1522
- Costa L, Giordano G (2007) Developmental neurotoxicity of polybrominated diphenyl ether (PBDE) flame retardants. *Neurotoxicology* 28:1047–1067
- Cranor CF (1993) Regulating toxic substances: a philosophy of science and the law. Oxford University Press, New York
- Cranor CF (1995) The social benefits of expedited risk assessment. *Risk Anal* 15:353–358
- Cranor CF (2008) Review of David Michaels. *Doubt is Prod Sci* 321:1296–1297
- Cranor CF (2011) Legally poisoned: how the law puts us at risk from toxicants. Harvard University Press, Cambridge, MA
- Dieter RR, Piepenbrink MS (2006) Perinatal immunotoxicity: why adult exposure assessment fails to predict risk. *Environ Health Perspect* 114:477–483
- Eriksson P, Viberg H, Johnsson N, Fredriksson A (December 7–10, 2009) Effects of perfluorinated compounds and brominated flame retardants on brain development and behavior in a rodent model In: Presentation at PPTOX II: role of environmental stressors in the developmental origins of disease, Miami, FL
- Faustman EM, Omenn GS (2001) Risk assessment. In: Klaassen C (ed) Casarett and Doull's toxicology, 6th edn. Pergamon, New York, pp 83–104
- Fimrite P (2009) Study: chemicals, pollutants found in newborns. *San Francisco Chronicle*, located at [www.sfgate.com](http://www.sfgate.com). Accessed 3 Dec 2009
- Grandjean P, Landrigan P (2006) Developmental neurotoxicity of industrial chemicals. *Lancet* 368: 2167–2178
- Guth J, Denison RA, Sass J (2007) Require comprehensive safety data for all chemicals. *New Solut* 17:233–258
- Hass U, Scholze M, Christiansen S, Dalgaard M, Vinggaard AM, Axelstad M, Metzdorff SB, Kortenkamp A (2007) Combined exposure to anti-androgens exacerbates disruption of sexual differentiation in the rat. *Environ Health Perspect* 115(Suppl 1):122–128
- Hattis D, Barlow (1996) Human interindividual variability in cancer risks: technical and management challenges. *Health Ecol Risk Assess* 2:194–220
- Heindel JJ (2008) Animal models for probing the developmental basis of disease and dysfunction paradigm. *Basic Clin Pharmacol Toxicol* 102:76–81
- Heinzow BGJ (2009) Endocrine disruptors in human breast milk and the health-related issues of breastfeeding. In: Shaw I (ed) Endocrine-disrupting chemicals in food. Woodhead Publishing, Cambridge, pp 322–355
- Jirtle RL, Skinner MK (2007) Environmental epigenomics and disease susceptibility. *Nat Rev* 8:253–262
- Kortenkamp A (2007) Ten years of mixing cocktails: a review of combination effects of endocrine-disrupting chemicals. *Environ Health Perspect* 115(Suppl 1):98–105
- Landrigan P (August 4, 2009) What's getting into our children? *New York Times* located at [www.nytimes.com](http://www.nytimes.com). Accessed 5 Aug 2009
- Lanphear BP (August, 2005) Origins and evolution of children's environmental health. In: Goehl TJ (ed) Essays on the future of environmental health research: a tribute to Kenneth Olden, special issue. Environmental Health Perspectives/National Institute of Environmental Health Sciences, Research Triangle Park, pp 24–31
- Merrill R (2004) FDA regulatory requirements as tort standards. *J Law Policy* 12:549–558
- Michaels D (2008) Doubt is their product: how industry's assault on science threatens your health. Oxford University Press, New York
- National Research Council (NRC) (1983) Risk assessment in the federal government: managing the process. National Academy Press, Washington, DC
- National Research Council (NRC) (1984) Toxicity testing: strategies to determine needs and priorities. National Academy Press, Washington, DC
- National Research Council (NRC) (2006) Science and judgment in risk assessment. National Academy Press, Washington, DC
- National Research Council (NRC) (2008) Toxicity testing for assessment of environmental agents: interim report. National Academy Press, Washington, DC
- Navas-Acien A, Guallar E, Silbergeld EK, Rothenberg SJ (2007) Lead exposure and cardiovascular disease—a systematic review. *Environ Health Perspect* 115:472–482
- Needham LL (2007) Personal communication, Faroe Islands
- Rescher N (1983) Risk: a philosophic introduction to the theory of risk evaluation and management. University Press of America, Washington, DC

- Santone KS, Powis G (1991) Mechanism of and tests for injuries. In: Hayes WJ Jr, Laws ER Jr (eds) *Handbook of pesticide toxicology*. Harcourt Brace Jovanovich, New York, pp 169–214
- Sexton K, Needham LL, Perkle JL (2004) Human biomonitoring of environmental chemicals: measuring chemicals in human tissues is the ‘gold standard’ for assessing people’s exposure to pollution. *Am Sci* 92:38–45
- Silbergeld EK, Weaver VM (2007) Exposures to metals: are we protecting the workers? *Occup Environ Med* 64:141–142
- State of Washington, Department of Ecology (2005) Dirt alert: Tacoma Smelter Plume, located at [www.ecy.wa.gov/programs/tcp/sites/tacoma\\_smelter/ts\\_hp.htm](http://www.ecy.wa.gov/programs/tcp/sites/tacoma_smelter/ts_hp.htm). Accessed May 4 2009
- U.S. Congress, Office of Technology Assessment (OTA) (1987) Identifying and regulating carcinogens. U. S. Government Printing Office, Washington, DC
- U.S. Department of Health and Human Services, Centers for Disease Control and Prevention (CDC) (2005) Third national report on human exposure to environmental chemicals, located at [www.cdc.gov](http://www.cdc.gov). Accessed 20 Aug 2008
- U.S. Department of Health and Human Services, Centers for Disease Control and Prevention (CDC) (2009) Fourth national report on human exposure to environmental chemicals, located at [www.cdc.gov](http://www.cdc.gov). Accessed 20 Dec 2009
- Wagner W (2008) Using competition-based regulation to bridge the toxics data gap. *Ind Law J* 83:629–659
- Wigle DT, Lanphear BP (2005) Human health risks from low-level environmental exposures. *PLoS Med* 2:1–3, located at [www.plosmedicine.org](http://www.plosmedicine.org). Accessed 5 Aug 2009
- Woodruff TJ, Zeise L, Axelrad DA, Guyton KZ et al (2008) Meeting report: moving upstream—evaluating adverse upstream end points for improved risk assessment and decision-making. *Environ Health Perspect* 16:1568–1575

# 32 Moral Emotions as Guide to Acceptable Risk

Sabine Roeser

Delft University of Technology, Delft, The Netherlands

University of Twente, Enschede, The Netherlands

<b>Introduction .....</b>	<b>820</b>
<b>Historical Developments .....</b>	<b>821</b>
Risk Perception .....	821
The Affect Heuristic .....	822
<b>Current Developments .....</b>	<b>822</b>
Dual Process Theory .....	823
Against DPT .....	823
Emotional Reflection and Correction .....	824
<b>Risk Emotions in Practice .....</b>	<b>826</b>
Emotions of Agents in Risk Society .....	826
Risk Emotions of the Public .....	826
Risk Emotions of Experts .....	826
The Role of Risk Policy Makers .....	827
Emotions and Risky Technologies .....	827
Emotions and Nuclear Energy .....	827
Emotions and Climate Change .....	827
Emotions and Risk Politics .....	828
<b>Further Research .....</b>	<b>829</b>
Financial Risks .....	829
Risk from Terrorism .....	829
<b>Conclusion .....</b>	<b>829</b>

**Abstract:** Risks arising from technologies raise important ethical issues for people living in the twenty-first century. Although technologies such as nanotechnology, biotechnology, ICT, and nuclear energy can improve human well-being, they may also convey risks due to, for example, accidents and pollution. As a consequence of such side effects, technologies can trigger emotions, including fear and indignation, which often leads to conflicts between experts and laypeople. Emotions are generally seen to be a disturbing factor in debates about risky technologies as they are taken to be irrational and immune to factual information. This chapter reviews the psychological literature that seems to support this idea. It then presents an alternative account according to which this is due to a wrong understanding of emotions. Emotions can be a source of practical rationality. Emotions such as fear, sympathy, and compassion help to grasp morally salient features of risky technologies, such as fairness, justice, equity, and autonomy that get overlooked in conventional, technocratic approaches to risk. Emotions should be taken seriously in debates about risky technologies. This will lead to a more balanced debate in which all parties are taken seriously, which increases the chances to be willing to listen to each other and give and take. This is needed in order to come to well-grounded policies on how to deal with risky technologies. The chapter discusses various recent examples of hotly debated risky technologies and how an alternative approach of emotions can help to improve debates about the moral acceptability of these technologies. The chapter ends with suggestions for future research in the areas of financial risks and security risks.

## Introduction

---

Risky technologies often give rise to intensive public debates. Examples of risks that spark heated and emotional debates are cloning, GM-foods, vaccination programs, carbon capture and storage, and nuclear energy. While large parts of the public are often afraid of possible unwanted consequences of such technologies, experts typically emphasize that the risks are negligible. They often accuse the public of being emotional, irrational, and wary to objective information. Policy makers usually respond to this gap between experts and public in either of two ways: by neglecting the emotional concerns of the public in favor of the experts or by accepting the emotions of the public as an inevitable fact and as a reason to prohibit a controversial technological development. Both of these responses are grounded on the assumption that the emotions of the public are irrational and block the possibility of a genuine debate. However, the assumption that emotions are irrational is far from obvious. To the contrary, many contemporary emotion scholars challenge the conventional dichotomy between reason and emotion. They argue that emotions are a form or source of practical rationality. This chapter argues that this alternative view of emotions can lead to a different understanding of emotional responses to risk. Risk emotions (i.e., emotions evoked by or related to risk or risk perception) can draw attention to morally salient aspects of risks that would otherwise escape our view. This alternative approach can shed new light on various controversial debates about risky technologies by showing the reasonableness of risk emotions. In addition, it can provide for a new approach on how to address emotions in debates about risky technologies. By taking the emotions of the public seriously, the gap between experts and laypeople can eventually be overcome, leading to more fruitful discussions and decision making.

## Historical Developments

---

Research into public risk perceptions started in the late 1960s and early 1970s with the rise of empirical decision theory. The initial focus was not so much on emotions, but on the way people make judgments under risk and uncertainty. It turned out that the risk judgments of people deviate substantially from the then academically dominant approach of rational decision theory, which was based on formal methods (see Part 3: “Decision Theory and Risk” of this handbook). Not only laypeople, but also experts, turned out to make decisions in ways that deviated from these strict rules, and to have problems processing statistical information (Tversky and Kahneman 1974, also see Gigerenzer 2002). This gave rise to a whole industry of investigations into the biases to which people are prone in risk judgments, under the header “heuristics and biases.” This research would eventually result in a Nobel Prize in economics for Daniel Kahneman in 2002.

## Risk Perception

---

Since the 1970s, Paul Slovic and his colleagues have conducted numerous psychometric studies into the risk perceptions of laypeople. This research began with the assumption that in so far as risk perceptions deviate from rational decision theory, they are biases. However, eventually Slovic started to develop an alternative hypothesis, namely, that it was possible that laypeople not so much have a *wrong* perception of risk, but rather a *different* perception of risk than experts. Maybe there was something to be learned from laypeople’s risk perceptions (Slovic 2000, p. 191). This hypothesis was supported by the finding that if asked to judge annual fatalities due to certain activities or technologies, laypeople’s estimates came close to those of experts. However, when asked to judge the *risks* of a certain activity or technology, laypeople’s estimates differed significantly from those of experts. Experts define risk as the probability of an unwanted effect, and most commonly, as annual fatality, so they perceive the two notions as by and large the same. However, apparently, for laypeople, these are different notions. They seem to have different connotations with the notion of risk that go beyond annual fatalities (Slovic 2000, pp. 113, 114). Slovic and his colleagues then started to conduct studies with which they tried to disentangle which additional considerations played a role in laypeople’s risk perceptions. They eventually developed a list of 18 additional considerations, including a fair distribution of risks and benefits, voluntariness, available alternatives, and catastrophic versus chronic risks (Slovic 2000, p. 86).

The question remains whether these considerations are reasonable concerns that should be included in risk assessments. The answer by sociologists and philosophers of risk to this question is positive. Whether a risk is acceptable is not just a matter of quantitative information but also involves important ethical considerations (see [Chap. 3, The Concepts of Risk and Safety](#), by Möller in this handbook) (cf. Krimsky and Golding 1992; Shrader-Frechette 1991; Hansson 2004). In the literature on ethical aspects of risk, the same considerations are brought forward as the ones that play a role in risk perceptions of laypeople.

Technocratic approaches to risk are based on the definition of risk as the probability of an unwanted effect and cost-benefit or risk-benefit analysis. Cost-benefit analysis resembles utilitarian theories in ethics, which state that we should maximize aggregate benefits or minimize unwanted outcomes. However, such approaches are subject to severe criticism in moral philosophy. Common objections against utilitarianism are that it ignores issues of fair

distribution, justice, autonomy, and motives. The same objections can be raised against cost-benefit analysis (Asveld and Roeser 2009). It is a morally important consideration how risks and benefits are distributed within a society (fairness, equality) (see also [Chap. 36, What Is a Fair Distribution of Risk?](#), by Hayenjelm in this handbook). Risks that are imposed against people's will are morally questionable (autonomy, cf. Asveld 2007). It is morally important whether a risk is due to intentional actions, negligence, or has occurred despite responsible conduct (motives) (see also [Chap. 33, Risk and Virtue Ethics](#), by Ross and Athanassoulis, and [Chap. 35, Risk and Responsibility](#), by Van de Poel and Nihlén Fahlquist in this handbook). A one-shot, catastrophic risk can be morally more problematic than a chronic, relatively small risk, even though the respective products of probability and effect might be similar. This is because in the case of a chronic risk, such as traffic risks, there are opportunities to improve outcomes, whereas in the case of a catastrophic risk, such as a nuclear meltdown, once it manifests itself, it can prove impossible to stop it, and the consequences can be disastrous for generations to come (Roeser 2006, 2007).

Hence, interestingly, laypeople, psychologists, social scientists, and philosophers share many of the same concerns when it comes to the moral acceptability of risk.

## The Affect Heuristic

---

Recently, another aspect of laypeople's risk perceptions has been investigated, namely, emotions. Melissa Finucane, Paul Slovic, and other empirical scholars have started to study the role of emotions, feelings, or affect in risk perception (cf. for example Alhakami and Slovic 1994; Finucane et al. 2000; Loewenstein et al. 2001; Slovic et al. 2002, 2004, and Slovic 2010). They have coined the terms "the affect heuristic" or "risk as feeling" to describe these perceptions (see [Chap. 26, The Role of Feelings in Perceived Risk](#), by Finucane in this handbook) (several journals have devoted special issues on this topic: *Risk Management* 2008, no. 3; *The Journal of Risk Research* 2006, no. 2). It turns out that emotions such as dread or fear significantly influence laypeople's risk perceptions. Some scholars see this as a reason to resist laypeople's risk perceptions, as they take emotions to be a disturbing factor for risk perception. Cass Sunstein (2005) emphasizes that emotions lead to what he calls "probability neglect." He proposes to use quantitative methods such as cost-benefit or risk-benefit analysis in order to come to a rational evaluation of risks. Paul Slovic thinks that "risk as feeling" should be corrected by "risk as analysis" (Slovic et al. 2004, p. 320). Others argue that we should respect the emotions of laypeople because we live in a democracy (Loewenstein et al. 2001), or for instrumental reasons, in order to create support for a technology (De Hollander and Hanemaaijer 2003). Hence, it seems like risk emotions constitute the following puzzle: the fact that emotions play an important role in laypeople's risk perceptions threatens to undermine the earlier claims about the broader risk rationality of laypeople. However, Slovic and his colleagues point out that emotions show us what we value (Slovic et al. 2004). It is this line of argument that this chapter explores in more detail in what follows.

## Current Developments

---

The idea that reason and emotion are diametrically opposed is deeply ingrained in our cultural and intellectual heritage, so much so that it is taken for granted and rarely questioned.

The same dichotomy can also been seen in the academic literature on risk and emotion. However, over the last decades, emotion-scholars have challenged the dichotomy between reason and emotion. Many leading philosophers and psychologists who study emotions argue that we need emotions in order to be practically rational. This idea can shed new light on the study of risk and emotion, which will be discussed in this section. This section will first discuss Dual Process Theory, the dominant approach to risk emotions. It will then discuss objections to this approach and present an alternative view.

## Dual Process Theory

---

The dominant approach in research into risk emotions is Dual Process Theory (DPT; cf. Epstein 1994; Sloman 1996; 2002; Stanovich and West 2002). According to DPT, people apprehend reality in two distinct ways: system 1 is emotional, affective, intuitive, spontaneous, and evolutionarily prior. System 2 is rational, analytical, computational, and occurred later in our evolution. System 1 helps us to navigate smoothly through a complex world, but it is not reliable, it provides us with heuristics, but also biases (cf. Gilovich et al. 2002). If we want to have reliable knowledge, we have to use system 2, but it takes more time and effort.

Neurological research by Joshua Greene is meant to support the framework of DPT. People who make utilitarian, cost-benefit moral judgments use rational parts of their brain; people who make deontological, respect-for-persons judgments use emotional parts of their brain (Greene 2003, 2007; Greene and Haidt 2002). Greene and also Peter Singer (2005) argue that this shows that utilitarian judgments are superior to deontological judgments, as the source of utilitarian judgments is superior, namely, reason rather than emotion.

DPT reflects the common dichotomy between reason and emotion: emotions are spontaneous gut reactions, but highly unreliable, reason is the ultimate source of objective knowledge, but it comes with the price of requiring more effort. This approach is commonly adopted by various scholars who study risk and emotion (e.g., Slovic 2010; Sunstein 2005). But the question is whether this is justified.

## Against DPT

---

There are developments in emotion research that cast serious doubt on DPT. Emotions are not contrary to knowledge and rationality, rather, they are a specific form of knowledge and rationality. Many contemporary emotion-scholars see emotions as a source of practical knowledge and rationality.

Groundbreaking research by Antonio Damasio (1994) has shown that people with a specific brain defect (to their amygdala in the prefrontal cortex; see [Chap. 27, Emotion, Warnings, and the Ethics of Risk Communication](#), by Buck and Ferrer in this handbook) have two problems: (1) They do not feel emotions anymore. (2) They cannot make concrete practical and concrete moral judgments anymore. This specific brain defect does not impair abstract rationality; people score equally high on IQ tests as they did before the injury or illness that caused the damage. These patients also still know abstractly that one should not lie, steal, or kill, etc. Their abstract moral knowledge and their abstract rationality are still intact. However, in concrete circumstances, these people do not know how to behave. They were initially virtuous, pleasurable people, but due to the brain defect, they changed into rude

people who act without consideration for others and cannot make concrete moral judgments. Their risk behavior is also affected. Damasio and his colleagues have developed the so-called Iowa-gambling task: an experiment in which people gamble in a lab setting. Where people without amygdala defects fall within a normal range of risk seekingness and risk aversion, amygdala patients have no risk inhibitions. They are willing to take major risks that normal people find unacceptable. Apparently, our emotions prevent us from taking outrageous risks, and they are necessary for making concrete moral judgments.

Other emotion scholars emphasize that emotions are not contrary to cognition but involve cognitive aspects (philosophers: e.g., de Sousa 1987; Greenspan 1988; Solomon 1993; Blum 1994; Little 1995, Stocker and Hegemann 1996, Goldie 2000; Ben Ze'ev 2000, psychologists: e.g., Scherer 1984; Frijda 1987; Lazarus 1991; Damasio 1994). Some scholars think that cognitions precede feelings which together constitute emotions (Reid 1969[1788]). Others propose the opposite model: Emotions are constituted by feelings that give rise to cognitions (Zajonc 1980; Haidt 2001). There are also scholars who argue that emotions are constituted by affective and cognitive aspects that cannot be pulled apart; they are two sides of the same coin (Zagzebski 2003; Roberts 2003; Roeser 2011a). Take the emotion of guilt. Experiencing this emotion involves feeling the “pangs of guilt.” Without the “pangs,” it is not genuine guilt. But it also means holding the belief that one did something wrong. The feeling aspect and the cognitive aspect of emotions go hand in hand. Emotions make us aware of moral saliences that would otherwise escape our attention (Little 1995; Blum 1994).

These insights can also shed new light on the research by Joshua Greene mentioned in the previous section. The fact that deontological judgments involve emotions does not undermine their status. Rather, this points to the limitations of utilitarianism and cost-benefit analysis, which is the predominant approach in conventional risk analysis. Making decisions based on utilitarian reasoning might sometimes be inevitable, but there are situations in which respect for persons should be the guiding line, for example in order to avoid deliberately sacrificing people to provide for a benefit for others (Roeser 2010a).

Based on this alternative understanding of emotions, we can say that moral emotions are needed in order to grasp moral aspects of risk, such as justice, fairness, and autonomy – aspects that cannot be captured by purely quantitative approaches such as cost-benefit analysis (Roeser 2006). Hence, rather than constituting a *puzzle* (see section [The Affect Heuristic](#)), emotions are the *explanation* why laypeople have a broader, ethically more adequate understanding of risk than experts: Because their risk perceptions involve emotions, they are more sensitive to the moral aspects of risk than the experts who mainly rely on quantitative methods (cf. Kahan 2008; Roeser 2010b).

## Emotional Reflection and Correction

The claim that emotions are necessary for moral judgments about risk does not entail that they are infallible. As with all sources of knowledge, emotions can misguide us. But whereas we can use glasses or contact lenses to correct our imperfect vision, there are not such obvious tools to correct our emotions. However, emotions can themselves have critical potential (Lacewing 2005). Sympathy, empathy, and compassion allow us to take on other points of view and critically reflect on our initial emotional responses. We can even train our reflective emotional capacities through works of fiction (Nussbaum 2001).

These ideas show how to correct misguided *moral* emotions. However, there are also emotions that are misguided about *factual* information, which may be especially poignant in the case of risk-emotions, because the information is complex and inherently uncertain.

In addition, there are ambiguous risk-emotions: emotions that can point out important moral considerations, but that can also be notoriously misleading. Prime candidates are fear and disgust. The question arises how we can distinguish between those forms of, for example, fear and disgust that point out moral saliences versus those that are based on stereotypes and phobias, and how to distinguish between them in political debates about risks.

The most visible and controversial emotion that is triggered by technological risks is fear (or worry or dread; cf. Slovic 2000). Ethical objections to new technologies such as cloning, human–animal hybrids, cyborgs, or brain implants are often linked to reactions of disgust. Where the alternative framework of emotions presented in this chapter can rather easily establish why moral emotions such as sympathy, empathy, and indignation should play an important role in political debates about the moral acceptability of risk, fear and disgust are more complicated. Fear and disgust are less clearly focused on moral aspects of risk, they can also be responses to perceived threats that might be based on wrong factual information. Fear and disgust might just reflect our unfounded prejudices and phobias, such as the fear of flying. Even in the light of contrary moral or factual evidence, we might still feel fear or disgust (cf. Sunstein 2005 for the irrationality of fear; cf. Haidt and Graham 2007 for the irrationality of disgust).

On the other hand, there are situations in which fear and disgust enable us to be aware of morally salient features. Interestingly, nanotechnology gives rise to greater worries within the scientific community than among the public (Scheufele et al. 2007). Given the newness of nanotechnology, we can assume that scientists are more knowledgeable than the public about nanotechnology and its concomitant risks. Apparently, their fears can be attributed to a rational understanding of the risks involved in nanotechnology. Fear can point to a source of danger to our well-being (Green 1992; Roberts 2003; Roeser 2009). In a similar way, disgust and the “uncanny” feelings we have concerning, for example, clones, cyborgs, human–animal hybrids and people with brain implants can point to our unclear moral responsibilities to them and the worry that they might develop in an unforeseen way. These are ethical concerns that need to be addressed in developing and dealing with such new technologies, and disgust can enable us to detect morally salient issues (cf. Miller 1997; Kahan 2000 on the rationality of disgust). Fear and disgust can be warning signs, making us aware of the moral values involved in new technologies. In so far as fear and disgust can sustain reflection (which can itself be an emotional process, cf. Lacewing 2005 and Roeser 2010c), they should inform our judgments.

Misguided risk emotions that are geared toward factual aspects of risk should be corrected by factual data. This is complicated by the fact that some risk-emotions function like stereotypes or phobias and can even be immune to factual information. Take the fear of flying; this emotion does not easily disappear in the light of evidence about the safety of plane travel. Factual information has to be presented in an emotionally accessible way in order to be able to correct misguided risk-emotions that are directed at factual aspects of risk (see [Chap. 27, Emotion, Warnings, and the Ethics of Risk Communication](#), by Buck and Ferrer in this handbook). One strategy might be to point out the benefits of a technology in cases where people focus on small risks.

However, it should be noted that not all biases in risk theory that are currently attributed to emotions are really based on emotions (Roeser 2010c). This might again be due to the

presupposition of DPT that irrational perceptions must be due to system 1, hence, emotion, but as argued before, this is an unwarranted claim (cf. Roeser 2009).

## Risk Emotions in Practice

---

This section discusses what the alternative approach to risk emotions means for the most important agents in risk society. Accordingly, it will discuss several areas of technology to illustrate how the alternative approach to risk emotions can shed new light on understanding the responses to these areas of technology. It will end with an alternative model for emotions in risk politics.

### Emotions of Agents in Risk Society

---

Research on risk emotions can focus on three main groups, i.e., the public, the experts, and policy makers. Each group has its own concerns, emotions, and moral considerations, but is also prone to potential biases, which can affect debates about risky technologies.

### Risk Emotions of the Public

Research on risk emotions mainly focuses on the public. In debates about risky technologies, public emotions are often the most visible and contrasted with the supposedly rational stance of experts. However, it is not clear that experts are free of emotions, nor is it clear that this is a bad thing, an idea that will be explored in the following section. In addition, as elaborated above, the fact that the public is emotional about risks might be the reason that they are capable of taking on a broader perspective on risk than the technocratic stance of experts and policy makers.

### Risk Emotions of Experts

It might be thought that experts take a purely rational, detached stance to risky technologies. However, scientists can be deeply emotionally involved with the research and technologies they develop (cf. McAllister 2005). As previously stated, experts are more worried about nanotechnology than the public (Scheufele et al. 2007). Experts are more knowledgeable about the scientific facts than laypeople, but that can also lead to increased moral concern and worry. Arguably, experts should take these worries and concerns seriously, which should lead to additional precautions (Roeser 2011c).

The emotions of experts can also be potential biases, for example, due to enthusiasm about their technologies or due to self-interested concerns, such as pressure on securing funding, positions, and prestige. Experts can control for these potential biases by also considering themselves as part of the public and trying to empathize with the point of view of potential victims of their technologies (cf. Van der Burg and Van Gorp 2005 who make this point based on an argument from virtue-ethics, which is extended to emotions in Roeser 2011c).

## The Role of Risk Policy Makers

Risk policy makers should ideally mediate between the insights of experts and the concerns of the public. However, in practice, there can be potential conflicts of interest that might be reinforced by emotions. For example, experts from industry lobby and often have close ties with the government when it comes to large infrastructural high-tech projects. Careers of policy makers can be at stake, which can lead to potential self-interested emotional biases. On the other hand, these same emotions can force politicians to follow the predominant views of the electorate. A virtuous policy maker should be someone who can balance the various considerations and emotions and take a wider perspective, based on feelings of responsibility and care for all members of society, in which risks and benefits of a technology and concomitant moral concerns are carefully balanced (cf. Nihlén Fahlquist 2009).

## Emotions and Risky Technologies

---

This section discusses two salient cases that show the practical applications of the theoretical framework sketched in the previous section, namely, nuclear energy and climate change.

### Emotions and Nuclear Energy

Due to the disaster at Fukushima, the debate about nuclear energy has taken an unexpected turn. In the last few years, there was a growing consensus that nuclear energy would be an important part of the solution to generate energy with decreased CO<sub>2</sub>-emissions. The probability of an accident was said to be negligible. However, now that an accident has occurred, nuclear energy has become controversial again and people argue that we should abandon it (cf. e.g., Macilwain 2011). Germany immediately shut down several nuclear reactors, and the German Green Party achieved unprecedented results in local elections due to its antinuclear position.

Despite this shift in focus, there seems to be one constant factor in the debate about nuclear energy: proponents call opponents badly informed, emotional, and irrational, using these notions more or less as synonyms. However, such rhetoric is denigrating and hinders a real debate about nuclear energy. In addition, it is simply wrong to equate emotions with irrationality, as they can be a source of practical rationality. A fruitful debate about nuclear energy should do justice to quantitative, empirical information as much as emotional, moral considerations. It should allow the public a genuine voice in which their emotions and concerns are appreciated, listened to, and discussed. By discussing the concerns underlying emotions, justified concerns can be distinguished from – morally or empirically – unjustified concerns (cf. Roeser 2011b).

### Emotions and Climate Change

Climate change is an urgent problem that will presumably affect the environment for generations to come, and it will also have effects on the health and way of life of present and future generations. Nevertheless, people seem to be unwilling to adapt their behavior. Several researchers who study the perceptions that people have of climate change have stated that

people lack a sense of urgency (Meijnders et al. 2001; Leiserowitz 2005, 2006; Lorenzoni et al. 2007; Lorenzoni and Pidgeon 2006).

Emotions are generally excluded from communication and political decision making about climate change, or they are used instrumentally to create support for a position. However, based on the account of emotions presented above, emotions can be seen as a necessary source of reflection and insight concerning the moral impact of climate change. In addition, emotional engagement also leads to a higher degree of motivation than a detached, rational stance on climate change (cf. Weber 2006). Hence, emotions might be the missing link in communication about climate change, in a twofold way: They lead us to more awareness of the problems and motivate people to do something about climate change (Roeser *under review*).

## Emotions and Risk Politics

---

Currently, many debates about risky technologies result in either of two pitfalls, namely what can be called the technocratic pitfall and the populist pitfall. In the technocratic pitfall, the debate is dominated by statistical, quantitative information, leaving no room for emotions and moral concerns. In the populist pitfall, the emotions and concerns of the public are taken for granted and seen as inevitable. If there is no public support, a risky technology cannot be implemented. Both pitfalls are due to the assumption that reason and emotion are distinct faculties. The technocratic pitfall favors “system 2,” reason and science. The populist pitfall favors “system 1,” emotion and gut reactions.

Sometimes both pitfalls occur in the same debate, for example, in the case of the debate about carbon capture and storage (CCS) in the Netherlands. Initially, CCS was to take place in Barendrecht. Experts told the concerned public that there were no risks involved. The concerns of the public were dismissed as being emotional and irrational. This can be seen as an instance of the technocratic pitfall. However, resistance was so strong that the initial plans were abandoned and rescheduled to the much less densely populated province of Groningen. However, the people there also rejected these plans from the start; why should they accept something that other people did not want in their backyard? Before there was even a genuine debate, politicians gave up this plan because there was apparently no “social support.” This can be seen as an instance of the populist pitfall: The will of the public is taken to be definite, with no attempt for a genuine dialogue on pros and cons.

However, the approach argued for in the previous sections offers an alternative to these pitfalls, based on a rejection of the dichotomy between reason and emotion. Emotions can be a source of moral reflection and deliberation. This allows us to avoid the technocratic and the populist pitfalls. Instead, we should endorse an “emotional deliberation approach” to risk. Emotions should play an explicit role in debates about risky technologies in the process of people discussing their underlying concerns. This approach may reveal genuine ethical concerns that should be taken seriously; it might also show biases and irrational emotions that can be addressed by information that is presented in an emotionally accessible way. This will enable a genuine dialogue. It will lead to morally better decisions, but as a side effect, it will also contribute to a better understanding between experts and laypeople. It is reasonable to assume that there will be a greater willingness to give and take if both parties feel that they are taken seriously. This procedure might seem more costly. However, it is likely to be more effective, and hence more fruitful in the long run. Currently, many debates about risky technologies result in

an even wider gap between proponents and opponents and in rejections of technologies that could be useful if introduced in a morally sound way. Genuinely including emotional concerns in debates about risky technologies can help overcome such predictable stalemates.

## Further Research

---

This section highlights a few possible topics for future research. The topics discussed in the previous sections are related to risky technologies. But the approach presented in this chapter might also shed new light on other debated risk issues in contemporary society, i.e., financial risks and the risks of terrorism (security risks).

### Financial Risks

---

The financial crisis that hit the world in 2008 has generally been attributed to the irresponsible conduct of the financial sector that was interested in short-term, self-interested gains only and a general disregard to long-term negative consequences for society. Based on the framework presented in this chapter, we can understand this as an example of how ethical considerations were sacrificed for supposedly objective, rational number crunching. Interestingly, the ideal of rationality in economics is self-interest. But from a philosophical point of view, the limited form of self-interest often celebrated in economics is frequently used as a justification to perform unethical actions (Powell 2010). In addition, selfishness can be a state fed by emotions. Such egoistic emotions should be corrected by moral emotions such as compassion and feelings of responsibility (cf. Frank 1988).

### Risk from Terrorism

---

Risks from terrorism (security risks) are often exaggerated and get disproportional attention as compared to other risks (cf. Sunstein 2005; De Graaf 2011). This phenomenon is usually blamed on emotions such as fear (Sunstein 2005). In response to terrorist risks, we can also observe the two pitfalls described earlier: either it is argued that we should ignore the fear of large parts of the population and respond to it with supposedly objective, factual, rational information (technocratic pitfall), or the societal unrest feeds into populist movements (populist pitfall). The approach presented in this chapter might open the way for an alternative approach, in which the emotions of the public are addressed in public debates but are open for revision. Such revisions could take place through factual information as much as through exercises in compassion with potential victims of measures that are supposed to limit the risks of terrorism but come at the price of xenophobia and disproportional measures that are insensitive to individual cases. Works of fiction might prove to be helpful for such exercises in compassion.

## Conclusion

---

This chapter has explored the role emotions do and can play in debates about risky technologies. Most authors who write on risk emotions see them as a threat to rational decision

making about risks. However, based on recent developments in emotion research, an alternative picture of risk emotions is possible. Risk emotions might be a necessary source of insights into morally salient aspects of risk. This view allows for fruitful insights on how to improve public debates about risk, and to overcome the gap between experts and laypeople that currently so often leads to a deadlock in discussions about risky technologies.

## Acknowledgments

Thanks to Jeff Powell for comments on a draft version of this chapter. Work on this chapter was supported by the Netherlands Organization for Scientific Research (NWO) under grant number 276-20-012.

## References

- Alhakami AS, Slovic P (1994) A psychological study of the inverse relationship between perceived risk and perceived benefit. *Risk Anal* 14:1085–1096
- Asveld L (2007) Autonomy and risk: criteria for international trade regimes. *J Glob Eth* 3(1):21–38
- Asveld L, Roeser S (eds) (2009) The ethics of technological risk. Earthscan, London
- Ben-Ze'ev A (2000) The subtlety of emotions. MIT Press, Cambridge, MA
- Blum LA (1994) Moral perception and particularity. Cambridge University Press, New York
- Damasio A (1994) Descartes' error. Putnam, New York
- De Graaf B (2011) Evaluating counterterrorist performance: a comparative study. Routledge/Francis & Taylor, Oxford/New York
- De Hollander AEM, Hanemaaijer AH (eds) (2003) Nuchter omgaan met risico's. RIVM, Bilthoven
- de Sousa R (1987) The rationality of emotions. MIT Press, Cambridge, MA
- Epstein S (1994) Integration of the cognitive and the psychodynamic unconscious. *Am Psychol* 49(8): 709–724
- Finucane M, Alhakami A, Slovic P, Johnson SM (2000) The affect heuristic in judgments of risks and benefits. *J Behav Decis Mak* 13:1–17
- Frank R (1988) Passions within reason: the strategic role of the emotions. W. W. Norton, New York
- Frijda N (1987) The emotions. Cambridge University Press, Cambridge
- Gigerenzer G (2002) Reckoning with risk. Penguin, London
- Gilovich T, Griffin D, Kahnemann D (eds) (2002) Heuristics and biases: the psychology of intuitive judgement. Cambridge University Press, Cambridge
- Goldie P (2000) The emotions. A philosophical exploration. Oxford University Press, Oxford
- Green OH (1992) The emotions. Kluwer, Dordrecht
- Greene JD (2003) From neural 'is' to moral 'ought': what are the moral implications of neuroscientific moral psychology? *Nat Rev Neurosci* 4:847–850
- Greene JD (2007) The secret joke of Kant's Soul. In: Sinnott-Armstrong W (ed) Moral psychology, vol 3, The neuroscience of morality: emotion, disease, and development. MIT Press, Cambridge, MA, pp 2–79
- Greene JD, Haidt J (2002) How (and where) does moral judgment work? *Trends Cogn Sci* 6:517–523
- Greenspan P (1988) Emotions and reasons: an inquiry into emotional justification. Routledge, New York/London
- Haidt J (2001) The emotional dog and its rational tail. A social intuitionist approach to moral judgment. *Psychol Rev* 108:814–834
- Haidt J, Graham J (2007) When morality opposes justice: conservatives have moral intuitions that liberals may not recognize. *Soc Justice Res* 20:98–116
- Hansson SO (2004) Philosophical perspectives on risk. *Techné* 8:10–35
- Kahan DM (2000) The progressive appropriation of disgust. In: Bandes S (ed) The passions of law. NYU Press, New York
- Kahan DM (2008) Two conceptions of emotion in risk regulation. *U Penn Law Rev* 156:741–766
- Krimsky S, Golding D (eds) (1992) Social theories of risk. Praeger, Westport
- Lacewing M (2005) Emotional self-awareness and ethical deliberation. *Ratio* 18:65–81
- Lazarus R (1991) Emotion and adaptation. Oxford University Press, New York
- Leiserowitz A (2005) American risk perceptions: is climate change dangerous? *Risk Anal* 25:1433–1442

- Leiserowitz A (2006) Climate change risk perception and policy preferences: the role of affect, imagery, and values. *Clim Chang* 77:45–72
- Little MO (1995) Seeing and caring: the role of affect in feminist moral epistemology. *Hypatia* 10:117–137
- Loewenstein GF, Weber EU, Hsee CK, Welch N (2001) Risk as feelings. *Psychol Bull* 127:267–286
- Lorenzoni I, Pidgeon NF (2006) Public views on climate change: European and USA perspectives. *Climatic Change* 77:73–95
- Lorenzoni I, Nicholson-Cole S, Whitmarsh L (2007) Barriers perceived to engaging with climate change among the UK public and their policy implications. *Glob Environ Chang* 17:445–459
- Macilwain C (2011) Concerns over nuclear energy are legitimate. *Nature* 471:549
- McAllister JW (2005) Emotion, rationality, and decision making in science. In: Hájek P, Valdés-Villanueva L, Westerståhl D (eds) Logic, methodology and philosophy of science: proceedings of the twelfth international congress. King's College Publications, London, pp 559–576
- Meijnders AL, Midden CJH, Wilke HAM (2001) Role of negative emotion in communication about CO<sub>2</sub> risks. *Risk Anal* 21:955–966
- Miller WI (1997) The anatomy of disgust. Harvard University Press, Cambridge
- Nihlén Fahlquist J (2009) The problem of many hands and responsibility as the virtue of care. In: Managing in critical times—philosophical responses to organisational turbulence proceedings
- Nussbaum M (2001) Upheavals of thought. Cambridge University Press, Cambridge
- Powell J (2010) The limits of economic self-interest. PhD thesis, Tilburg University
- Reid T (1969[1788]) Essays on the active powers of the human mind. Introduction by Baruch Brody. MIT Press, Cambridge, MA/London
- Roberts RC (2003) Emotions. An essay in aid of moral psychology. Cambridge University Press, Cambridge
- Roeser S (2006) The role of emotions in judging the moral acceptability of risks. *Saf Sci* 44:689–700
- Roeser S (2007) Ethical intuitions about risks. *Saf Sci Monit* 11:1–30
- Roeser S (2009) The relation between cognition and affect in moral judgments about risk. In: Asveld L, Roeser S (eds) The ethics of technological risks. Earthscan, London, pp 182–201
- Roeser S (2010a) Intuitions, emotions and gut feelings in decisions about risks: towards a different interpretation of “neuroethics”. *J Risk Res* 13:175–190
- Roeser S (ed) (2010b) Emotions and risky technologies. Springer, Dordrecht
- Roeser S (2010c) Emotional Reflection about Risks. In: Roeser S (ed) Emotions and risky technologies. Springer, Dordrecht
- Roeser S (2011a) Moral emotions and intuitions. Palgrave Macmillan, Basingstoke
- Roeser S (2011b) Nuclear energy risk and emotions. *Philos Tech* 24:197–201
- Roeser S (2011c) Emotional engineers: toward morally responsible engineering. *Sci Eng Ethics* (forthcoming)
- Roeser S (2011d), Risk communication, moral emotions and climate Change. *Risk Anal* (under review)
- Scherer KR (1984) On the nature and function of emotion: a component process approach. In: Scherer KR, Ekman P (eds) Approaches to emotion. Lawrence Erlbaum, Hillsdale/London, pp 293–317
- Scheufele DA, Corley EA, Dunwoody S, Shih T-J, Hillback E, Guston DH (2007) Scientists worry about some risks more than the public. *Nat Nanotechnol* 2:732–734
- Shrader-Frechette K (1991) Risk and rationality. University of California Press, Berkeley
- Singer P (2005) Ethics and intuitions. *J Ethics* 9:331–352
- Sloman SA (1996) The empirical case for two systems of reasoning. *Psychol Bull* 119:3–22
- Sloman SA (2002) Two systems of reasoning. In: Gilovich T, Griffin D, Kahneman D (eds) Heuristics and biases: the psychology of intuitive judgement. Cambridge University Press, Cambridge, pp 379–396
- Slovic P (2000) The perception of risk. Earthscan, London
- Slovic P (2010) The feeling of risk: new perspectives on risk perception. Earthscan, London
- Slovic P, Finucane M, Peters E, MacGregor DG (2002) The affect heuristic. In: Gilovich T, Griffin D, Kahnemann D (eds) Heuristics and biases: the psychology of intuitive judgement. Cambridge University Press, Cambridge, pp 397–420
- Slovic P, Finucane M, Peters E, MacGregor DG (2004) Risk as analysis and risk as feelings: some thoughts about affect, reason, risk, and rationality. *Risk Anal* 24:311–322
- Solomon R (1993) The passions: emotions and the meaning of life. Hackett, Indianapolis
- Stanovich KE, West RF (2002) Individual differences in reasoning: implications for the rationality debate? In: Gilovich T, Griffin DW, Kahneman D (eds) Heuristics and biases: the psychology of intuitive judgment. Cambridge University, New York, pp 421–440
- Stocker M, Hegemann E (1996) Valuing emotions. Cambridge University Press, Cambridge
- Sunstein CR (2005) Laws of fear. Cambridge University Press, Cambridge
- Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. *Science* 185:1124–1131

- Van der Burg S, Van Gorp A (2005) Understanding moral responsibility in the design of trailers. *Sci Eng Ethics* 11:235–256
- Weber EU (2006) Experience-based and description-based perceptions of long-term risk: why globalwarming does not scare us (yet). *Clim Chang* 77:103–120
- Zagzebski L (2003) Emotion and moral judgment. *Philos Phenomenol Res* 66:104–124
- Zajonc RB (1980) Feeling and thinking: preferences need no inferences. *Am Psychol* 35:151–175

# 33 Risk and Virtue Ethics

Allison Ross<sup>1</sup> · Nafsika Athanassoulis<sup>2</sup>

<sup>1</sup>London, UK

<sup>2</sup>Oswestry, UK

<i>Introduction</i> .....	834
<i>What Is Virtue Ethics?</i> .....	835
<i>The Case for a Virtue Ethics Approach to the Moral Evaluation of Risk</i> .....	837
<i>Character: The Basics</i> .....	840
<i>Habit, Education, and Moral Development</i> .....	843
<i>Decision Making</i> .....	845
<i>To Time Travel or Not to Time Travel?</i> .....	850
<i>Conclusion</i> .....	854
<i>Further Research</i> .....	854
<i>Notes</i> .....	855

**Abstract:** In this chapter, we explain the nature of virtue ethics, differentiating it from competing moral theories – consequentialism or deontology – and arguing that it is superior to both when it comes to the moral assessment of risk. We explore in detail what a virtue ethics approach to the moral evaluation of risk taking would involve, focusing particularly upon the role played by character in such assessments. Our main argument is that individual instances of risk taking are not isolated events, but part of a pattern of behavior on the part of the risk taker. We argue, furthermore, that this pattern does not arise as a result of arbitrary, automatic processes over which individual agents have no control. Rather, risk-related behavior patterns are the product of a complex set of settled dispositions that constitute character. We argue that character dispositions are developed over time through education, which involves habituation, active reflection, and reflective self-modification. They bring together the influences of desire, emotion, and thought to provide explanations of actions and decisions, which are multi-dimensional and profoundly sensitive to the particularity of individual risk-involving actions and choices. Risk taking is both a necessary part of human life and a source of moral vulnerability – it is very difficult to make good choices about risk and there are a lot of different ways in which our risky choices could prove to be morally inadequate. It is our contention that only virtue ethics with its emphasis on character – character development, and character vulnerability – provides us with a sufficiently rich vocabulary to (a) furnish satisfying explanations of the sensible moral judgments we make about risks and risk takers all the time and (b) facilitate effective rational reflection about commonsense moral evaluations of risk taking. We illustrate the value of the virtue ethics approach using a hypothetical time-travel experiment in which an agent must choose whether to take some very serious risks and/or whether to expose others to risk.

## Introduction

---

Imagine the following scenario. You are a researcher at the Uber-Tech Institute. A friend and colleague of yours, Bob, whose knowledge and skill you respect enormously, takes you aside and reveals that he has secretly been designing and building a time machine. The machine is now ready to be tested, but the process requires three participants – one to stay at the controls, another to enter the machine and travel in time, and a third to be a back-up time traveler, in case the first one needs to be rescued. Bob would like you to help him test the machine. He has already recruited his super-bright 13-year-old sister and he offers to let you choose which of the three roles you would like to play in the test. Before you decide, however, he does warn you that time travel has never happened before so there is no way of knowing what effects it will have on the traveler's physical and mental health. In addition, the testing process would entail making small changes to the past/future, which have a small probability of adversely affecting the present for some unknown third parties in unpredictable ways. What would you decide to do?<sup>1</sup>

There are a number of decisions you would need to make in the above scenario: a decision about how much and whether the kinds of risk involved are of the sort and scale to which it is acceptable to subject yourself, a decision about whether and when it is right to condone and abet other people taking very serious risks with their own physical and mental well-being, and a decision about how much risk and what sorts of risk is acceptable to subject unknowing third parties to, etc. These are all complex and difficult decisions and it would be convenient if there was a formula for making good and right decisions about whether, when, and what to risk. It

will be the contention of this chapter that there is no such formula to be had. We will argue that the reason for this is that the sorts of decisions considered above are intrinsically moral decisions and that, while some moral theories do purport to offer formulae or rules for decision making about risk, to attempt to do so is misguided. In place of rules and guidelines, we recommend a virtue ethics approach to moral decision making. Virtue ethics focuses attention away from rules and upon the features of decision making itself, supposing that good decisions are made when the broad range of mental faculties involved function well and harmoniously. Such decisions will take into account and give appropriate weight to the context of choice, the role of the reasoner, historical and other constraints (e.g., rights) as well as potential consequences. However none of these criteria will alone act as a hallmark or determiner of morally good choices or have decision-independent priority over the others.

Virtue ethics draws upon the sort of approach to moral matters that was taken by the ancient Greeks and particularly upon the ethical works of Aristotle. This chapter begins with an introductory section, which presents some basic ideas in Aristotelian virtue ethics for readers who would benefit from some background to the theory. Readers who are familiar with virtue ethics may wish to skip to the section entitled: [❷ The Case for a Virtue Ethics Approach to the Moral Evaluation of Risk](#) which outlines the case for a virtue ethical approach to risk decision making. The section [❸ Character: The Basics](#) introduces the notion of moral character, which plays a central role in virtue ethics and accounts for the unique perspective to risk-taking decisions, which is afforded by a character-based theory. In the section [❹ Habit, Education, and Moral Development](#), we consider the long and gradual process of character development, while the section [❺ Decision Making](#) considers what is involved in making virtuous moral judgments. Links to how this distinctive virtue ethical perspective affects how we should reason about risk are made throughout the chapter, but the section [❻ To Time Travel or Not to Time travel?](#), in particular, returns to considering the time-traveling example introduced above in light of the claims we have made on behalf of a virtue ethical approach to making decisions about risk.

## What Is Virtue Ethics?

The modern revival of interest in virtue ethics began with the work of philosophers such as Elizabeth Anscombe, Bernard Williams, and Alasdair McIntyre. These authors raised a number of different ideas, many of them critical of the other two main alternative moral theories, deontology, and consequentialism, but if there is one thought which above all others captures the spirit of the discussion, it is this: modern virtue ethicists have redefined the kind of question we should ask about ethics. Ethical questions deal with practical matters and as such there is often a focus on specific problems. What should I do when faced with X? If I have to choose between Y and Z, which one should it be? What is the right thing for me to do *now*? Deontology and consequentialism give alternative, and in many ways incompatible, answers to these kinds of questions about practical issues. The advice of consequentialism is that decision making should be goal oriented and can be captured in one overriding rule, for example, do whatever maximizes the greatest consequences for the greatest number. The advice of deontology is that you may take any course of action, which is compatible with strict observance of the rights of others<sup>2</sup>. Both take it that the key question in ethics is “how ought I to act now?” Virtue ethics, on the other hand, holds that the rightness or wrongness of individual actions and choices,

whilst not insignificant, is not the key moral issue. Rather our attention ought to be focused upon questions concerning the sort of beings/persons we are (and will be) should we choose and behave in one way rather than another.

Virtue ethics, then, redefines the kind of question we should be asking in ethics. Instead of asking what is the right thing to do here, virtue ethics suggests we should be concerned with what kind of person we should become, and what kinds of lives we should live. Asking how one should live one's life gives ethical enquiry a different perspective. The object of the enquiry is now an entire life, a long project of self-reflection and self-development, which is both approached as a long-term commitment and judged as such. Good moral judgment is a product of this long process of reflection and development and getting a particular ethical call right on a particular occasion has as much to do with having a well-organized, appropriately responsive, and correctly structured/balanced character as it does with actual rights or consequences.

These ideas have their roots in Aristotle, so Aristotle's ethics is a good starting point for us as well. Aristotle begins his deliberations on ethics by reflecting upon the nature of his subject matter. He notes that ethics is a complex and varied subject, so when we ask questions about ethics we should expect our answers to be complex and varied too. It is no good attempting to capture a diverse and challenging subject in a simple and all-encompassing rule. A "one rule fits all" might sound like a good idea if we wish to dispense with ethical problems with the minimum amount of hassle and reflection, but such an approach is bound to fail as it cannot, by its very nature, capture the diversity of the subject at hand. So, we should expect to find complex, varied, and diverse answers to ethical questions and, clearly, such challenging answers are not going to be easy to come by (NE1098a). Virtue ethics then offers a radically different approach, both in the kind of question it asks, that is, how should I live my life, and in the kind of answer it expects to find, that is, a complex, varied, and imprecise answer that cannot be captured in an overriding rule.

Aristotle starts by noting that every action and pursuit aim at some good and that while some things are done for the sake of others, there must be something which is desired for its own sake, an ultimate goal for the sake of which everything else is done; that goal is *eudaimonia* (NE1094a). *Eudaimonia* is a challenging term to translate. "Happiness" is a popular translation, but it tends to suggest an ephemeral, transient feeling, which is easily affected by external factors and which has a specific target, for example, "I am happy today because I won at the riding competition" – this feeling is both generated by my win and dependent on it, as well as being likely to dissipate in time. "Contentment" is a slightly better translation in that it captures a sense of permanent and stable satisfaction with one's overall life, although it has connotations of passivity and resignation and there is nothing meek or weak about *eudaimonia*. Perhaps, "fulfilment" or "flourishing" is the best term, but it has to be understood within the Aristotelian context, which we explain in what follows.

*Eudaimonia* is not pleasure, for it is determined by the value of the activity which gives rise to it and we are looking for something, which is valuable in itself. Nor can it be honor, for that depends on those who confer it and we are looking for something which is truly characteristic of the person and not an external attribute easily given and taken away. Nor can it be wealth, for that is merely a means to other things and not an end in itself (NE1095a17-1096a15). Perhaps to discover what *eudaimonia* is, we need to consider our function or purpose. Aristotle observes that where a thing has a function, the good of that thing, that is, when we say that that thing is doing well, is to perform that function well. This idea is best understood by example. The knife has a function, which is to cut. The good of the knife, that is, when we say

that the knife is doing well, is to perform that function well, that is, to cut well. So, a good knife is a knife that cuts well (NE1096a25–33).

The same sort of argument, Aristotle suggests, can be applied to human beings. That is to say that in order to answer the question of how we should live our lives, we need to consider what kind of being human beings are and to discover what human life is for or what kinds of lives are characteristic of beings such as ourselves. Human beings, he thinks, have a function, so when we say that human beings are doing well it's because they are performing their function well. All we need to do now is to discover the function of human beings. To discover the function of human beings, we need to consider what is distinctive of human beings, what sets human beings apart from other beings, and is peculiar to us *qua* human beings. The answer is the ability to reason. The function of human beings is reason and the life which is distinctive of humans is the life in accordance with reason. If the function of human beings is reason, then the good human being is the human being who functions well, that is, reasons well, and reasoning well is the life of excellence for human beings (NE1097b21–1098a15, for a modern interpretation of these kinds of arguments from Aristotle, see Hursthouse 1999). Reason, according to Aristotle, takes particular forms depending upon what is being reasoned about; one of these forms is practical reasoning which occupies itself with questions of value and seeks to bring actions in accordance with values. Practical rationality is a key element of the complex psychology of action. So, part of what it takes to flourish is to exercise good practical rationality; the person who achieves this is called practically wise (*phronimos*). Aristotle famously describes two kinds of *eudaimon* life: the life of contemplation and the life of the *phronimos*. The former does not lack practical wisdom, but simply seldom encounters situations where it is necessary to deploy it; the latter is fully engaged in social and political life so that the exercise of practical wisdom is a major (perhaps primary) manifestation of its subject's capacity for reason (NE1102a). A person who is practically wise will also be virtuous. This follows because a practically wise person makes the best possible (all things considered) decisions about how to act and, in most cases, has the confidence and self-possession to enact them. *Eudaimonia*, then, consists in activity in accordance with reason, and leads to fulfilment and contentment with one's life and the highest activity in accordance with reason is virtue – moral goodness, moral excellence.

## **The Case for a Virtue Ethics Approach to the Moral Evaluation of Risk**

---

Actions involving risk have a particular nature, one that makes moral judgments about these actions particularly problematic. Actions involving risk are hostages to fortune, in that, by definition, the results of one's actions are not (at least not entirely in some cases) under one's control. Risky outcomes may or may not actualize, but whether they do so or not, is (at least partly) outside the control of the agent who instigated the risky actions<sup>3</sup>. This very nature of risky actions poses problems for moral responsibility. We hold people morally responsible and subject them to moral praise or blame because we fundamentally assume that they had control over their actions. This thought captures what is distinctive about judgments of moral responsibility and ties in with our understanding of agency. Agents are free beings who demonstrate their agency, that is, their wills, their choices, their reasoning, etc., in their actions, which is exactly why we hold them responsible for these actions. Where agency and control come apart, for example, in cases of people who are severely mentally ill or people who are not

yet fully formed moral agents, that is, children, we accept that there is limited scope for attributions of moral responsibility. This creates a problem for moral judgments of risky actions, for how are we to hold agents responsible for risky outcomes, which materialized due to the vagaries of luck rather than the effects of agency?

The nature of the problem is made clear by consideration of how inadequate some consequentialist theories are for assessing risk. Consequentialist theories evaluate the morality of actions based on the value of their consequences so that a morally good action is one that, for example, produces the greatest utility for the greatest number of people. However, it is exactly this focus on consequences that produces distorted results when applied to actions involving risk. Situations involving risk involve, of necessity, uncertainty; therefore the outcomes of one's actions will be uncertain. One possible response to this problem is to evaluate an action based on the actual consequences, those consequences that come about once the risk has actualized. However, this is clearly counter-intuitive in terms of moral responsibility. Consider the vice of recklessness, that is, the indifferent disregard for the consequences of one's actions, which becomes a lot more problematic when one's actions affect others. A reckless action is reckless because of the character trait it displays, rather than because of the actual outcomes it produces. Suppose an agent chooses to neglect his/her car's routine maintenance out of sheer boredom, then gets drunk out of sheer self-indulgence knowing he/she is likely to drive him/herself home and then speeds on his/her way home out of sheer extravagant enjoyment of fast driving. This fast, drunk driving in an unsafe car, is reckless because of its irresponsible disregard for the safety of other road users, the well-being of whom could have been better safeguarded had the driver in question been concerned enough to take some modest actions in that respect, for example, service his/her car, remain sober, and respect the speed limit. The moral judgment concerning the reprehensibility of the drunken driving concerns the character trait of recklessness that led the agent to behave in this way rather than the actual consequences of his/her action. For despite all his/her recklessness and the clear endangering of other people's welfare, our drunken driver could end up being lucky, that is, despite the high likelihood of him/her injuring someone, he/she could actually avoid this outcome. However, avoiding the consequences of one's recklessness does not make one any less responsible for it. That the driver was lucky and that the small chance of him/her getting home safely actualised does not make his/her risk taking any more acceptable.

One may respond here that we should not be concerned with actual consequences, but rather with expected consequences. For whatever the actual consequences of such reckless driving may be, the expected consequences of driving so recklessly are that someone is likely to be hurt and it is this aspect of the agent's action that we should hold him/her responsible for. However, estimating expected consequences in situations of risk can be problematic and even impossible in some cases (Hansson 1993). Furthermore, even if we were to set such problems of calculating expected outcomes aside, the impersonal calculation of consequences leaves no room for partial considerations, especially those that allocate greater weight to significant sacrifices by particular individuals (Hansson 2003), or those that differentiate between the bearers of different risks and benefits (Athanasoulis and Ross 2010). The first concern is that strict calculations of expected consequences do not do justice to the relative burdens born by different individuals. Consider the following example: suppose that as a result of an industrial accident at a nuclear power plant, there is a leakage of radioactive gasses. The leakage is relatively simple to stop in that it requires no particular expertise, anyone can enter the room to carry out the necessary operations; however, the concentration of radioactive gasses

in the room in question is so high that in stopping the leak, the person carrying out the repair will be exposed to high levels of radioactivity that have a 90% chance of killing him/her. Alternatively, the gases could be allowed to escape the building, rendering the room safe for the repair, but at the risk of exposing 1,000 people in the immediate vicinity to a risk of 0.001 of being killed by radioactive poisoning. Pure calculation of numbers suggests that the unfortunate security guard who happens to be in the vicinity of the leak should step into the room to carry out the repairs. However, this outcome, although it makes sense numerically, clashes with our sense of fairness with respect to the equitable distributions of the burdens of risk taking. Concentrating all the negative effects of the risk on one unfortunate individual seems a lot less fair than distributing a much smaller burden over larger numbers of people and some disadvantages for some people cannot be justified by the cumulative advantages conferred to a large number of other people.<sup>4</sup>

The second concern with consequentialist approaches to the moral evaluation of actions involving risk is a worry that the strict calculation of consequences does not allow room for differentiating between the bearers of risks and benefits. In some cases of risky action, the person or persons who run the risk of being harmed by the action may be different from the person or persons who run the risk of benefiting from the outcome and this feature makes these situations particularly problematic. Quite high risk of harm and even the risk of death may both be acceptable if one and the same individual stands to gain from the act. For example, a terminally ill patient may choose to take part in a Phase III trial of a promising, but untested, drug that also runs the risk of ending the patient's life prematurely. The background of the terminal illness for which all therapeutic options have been exhausted, coupled with the fact that the patient willingly risks only his/her *own* well being, make this a justifiable decision. However, the same cannot be said for actions which risk the welfare of others to benefit an individual. For example, an oil company may take significant risks by drilling in deep water with machinery which has been tested but not in the extreme conditions of the deep oceans. If the oil should spill during drilling, the impact upon the ocean and the inhabitants of nearby territories could be devastating (including possible death, illness, loss of livelihood, etc.). The oil company takes this risk in order to keep profit margins high and, if successful, stands to benefit by dominating the oil industry. Say that the risk of leakage is small and as it turns out, nothing bad happens. According to consequentialists, the small risk of leakage must be adjudged a case of acceptable risk taking. However, this does not seem to be quite right, the oil company is still unfairly putting people's lives and livelihoods at risk. Consequentialist approaches are not sensitive to these distinctions and leave much to be desired as they evaluate risk taking without accounting for the integrity of the persons making the decisions.

This sense of dissatisfaction with consequentialist approaches to risk is, however, useful because it points toward a different alternative. When we discussed recklessness above, we captured what was morally problematic about being indifferent to the risks to which one exposes people by referring to the *character trait* of recklessness. This notion of *character* gives us a different approach to the moral assessment of risk. Discussions of risk, strongly influenced by consequentialism, tend to focus on one-off, exceptional circumstances, dramatic choices that few individuals are ever unlucky enough to have to make. A character-based theory of risk would shift the focus from such high profile, but rare and therefore unrealistic, choices, to everyday concerns. If the focus is on one's character, then we have to examine the patterns in choices that people make, those choices that are reaffirmed over time, and those choices that express their deeply held values and beliefs which are hardly a matter for one-off exceptional

circumstances. The discussion of risk shifts from the exceptional to what is characteristic of individuals and this makes more sense of our common reactions to risk. The reckless person is not likely to be reckless in a one-off, extraordinary situation, but rather to exhibit this character trait of recklessness over time, consistently in different kinds of cases and with reliability. This is because recklessness is an attitude to risk, which is underpinned by a judgment about the relative unimportance of other people's well being. It is essentially characterized by a disregard for the welfare of others, an entrenched attitude that is displayed over time and in a variety of situations. A character-based theory of risk then would be more concerned with judgments of agents over time and over a variety of different situations, all of which illustrate the person's character, than with one-off, albeit spectacular, but extraordinary occurrences. A character-based theory of risk will be less about dramatic choices and more about the people involved in these decisions and how all their choices, over a variety of decisions, determine and illustrate who they are.

## Character: The Basics

---

The notion of "character" in virtue ethics refers to a specific and technical term. The word "character" comes from the Ancient Greek for carving, indicating a permanent and indestructible way of preserving something and this gives us a good indication of the use of the word nowadays. One's character is the set of stable, permanent, and well-entrenched dispositions to act in particular ways (Athanassoulis 2005, pp. 27–34). In the same way that a carving is a permanent and reliable mark, one's character is the collection of permanent and reliable dispositions, which characterize who one is. One's character leads one to act in a particular way, but these actions are also an expression of who one is, and one's behavior is explained by one's character. At the same time, when we speak of someone with no character, we tend to mean that that person has no strength of will, yields to the wishes of others easily, is overcome by temptation, and cannot be depended upon to act in a particular way consistently. So, the notion of character is essentially captured in the ideas of stability, reliability, predictability, and permanence (Kupperman 1991 esp. Part I).

Character dispositions are long-term features of agents, but they are not the *only* long-term features of agents, so it is important to provide a more substantial account of character, which can distinguish character from these other sorts of characteristics of agents. We cannot hope to offer a detailed comparison of character, identity, and personality here, but such an account can be found in Kupperman (1991, pp. 3–46). For our purposes, a preliminary narrowing of the definition of character can be made by specifying that it has to do with the ways we think and act, and that moral character concerns the way we think and act in moral situations. Moral character will be the focus of our discussion. Aristotle thinks of states of character (such as virtue or vice) as complex rather than simple states of mind. Virtue, for example, involves more than being in possession of a true or logically consistent belief about how one ought to act or being sentimentally generous to others. Indeed, what is required by Aristotle is a state in which the faculties of perception, motivation, thought, and reason seamlessly interact to bring about cogent appropriate action in individual cases and establish the long-term possession of a stable disposition to respond well to whatever situation is encountered. Consider his definition of virtue in the *Nicomachean Ethics*:

- ▶ Virtue, then, is a state of character concerned with choice, lying in a mean, i.e. the mean relative to us, this being determined by a rational principle, and by that principle by which the man of practical wisdom would determine it (NE 1107a).

For present purposes, what we are interested in extracting from this definition is what it tells us about character in general. What we can deduce from this definition is that states of character are intimately linked to action through choice. To possess character is to manifest it in the form of choices about how to act. In addition, this definition makes clear that choices and actions which manifest character are the product of the interaction of reaction, reason, and motivation. The way in which these relationships function is not something that is evident from choice or actions themselves, and it is for this reason that virtue or states of character are frequently described as dispositions; as here in the *Nicomachean Ethics*:

- ▶ There are three kinds of disposition [that can be constitutive of character], then, two of them vices, involving excess and deficiency respectively, and one a virtue, viz. the mean, and all are in a sense opposed to all; for the extreme states are contrary both to the intermediate state and to each other, and the intermediate to the extremes (NE 1108b).

So states of character are complex dispositions. Like all other dispositions, they will have the following characteristics:

- They are latent rather than manifest. We can only know of their existence through experience of the manifestations with which they are associated.
- They rely for their manifestation, although not for their existence, upon the obtaining of certain environmental conditions.

One's character is a state of being encompassing one's settled and stable dispositions that will, under normal circumstances, manifest themselves in action. That is, being kind will be the settled and stable disposition to display kindness in situations which call for a kind response. There is a clear connection then between one's character and the actions one chooses to perform, and conversely one's actions will manifest one's character.

We wish to illustrate this by returning to our time travel example from the beginning. Bob's choice of his young sister as one of the experiment participants who might be exposed to unknown and possibly serious risks, tells us something about his character, which, in turn, allows us to assess this decision to take risky action irrespective of the outcomes of the experiment. Bob's sister is both relatively young and therefore less likely to be able to weigh up correctly the serious risks involved in participating in such an experiment, and related to Bob with family bonds that may affect the voluntariness of her decision to participate. These concerns may lead one to worry about Bob's attempt to recruit his sister to the experiment and lead one to wonder whether Bob's judgment is clouded by any of the following:

- His devotion to and enthusiasm for the project of time travel
- His anxious state of mind
- A particular habitual pattern of interaction between him and his sister
- Pressure from his sponsor to "get on with it"

Bob's apparently reckless behavior toward his sister is indicative of his character, it appears to expose a simple disposition to endanger others for his own gain. We need to be careful however about inferring simple dispositions from this one instance of decision making. This

particular occasion may be a one-off, a situation of great temptation and difficulty, which has challenged Bob's otherwise sensible behavior. However, it may also be part of a pattern of such reckless actions that manifest itself over time. The real nature of this recklessness in Bob's character, cannot really be known without observing and understanding the behavior resulting from Bob's character over time. Observing Bob over time will help us determine whether he is habitually reckless – a vice – or prone to recklessness when faced with difficult or tempting situations – a case of weakness of will. In either case, the assessment of Bob's attitude to risk is not a matter of observing his one-off responses, one would need to get to know Bob's character over time before any such judgment could be made (Athanasoulis 2000).

Given this account of character, then, we would expect a character-based theory of risk to offer us not only a different answer to whether a risk should be taken in any one particular case, but also, more importantly, a different way of approaching the question. The focus of assessment turns from tangible consequences, probabilities, or individual intentions to produce consequences, toward the agent as a whole, toward her quality as a decision maker and therefore toward the quality of the way in which she lives. To illustrate the value of the “whole agent” approach to moral evaluation, let us consider an example, which is slightly less complex than Bob's. Consider the comparison of two mothers and their responses to the prospect of a potentially risky vaccination that is available for their children. Mother One informs herself by reading up on the disease, its prevalence, its side effects, possible treatments, efficacy of the vaccine, adverse effects of the vaccine, etc. Her research suggests that whilst the risk of the vaccination is small and unlikely compared with the risk of the disease, there is nevertheless reason to fear an adverse reaction. She considers her overall commitment to make decisions on behalf of her child based on the child's best interests and despite her natural fear for her child's well being, she reasons that not to vaccinate is to leave him vulnerable to horrible diseases, and that vaccinating also means contributing to public health. She decides that the risks involved in vaccination are worth taking in order to protect him from such vulnerability not purely because of the possible outcomes, but also because of the specific perspective she ought to adopt as the child's primary care giver. That is, being responsible for the welfare of another human being and having to make important healthcare decisions on their behalf, imposes an extra burden of caution. Risks and outcomes are then evaluated through the filter of this perspective of “deciding on behalf of a vulnerable other for whom I have special care responsibilities,” a perspective that changes the weight one allocates to possible dangers and benefits. Now lets say she vaccinates and her child has the rare and unlikely adverse reaction. Compare Mother One with Mother Two who makes an uninformed decision not to vaccinate, based primarily on the inconvenience of getting the vaccine. Fortunately, her child happens to experience little exposure to the disease concerned and benefits from herd immunity and so no harm results. Which mother in this example acts well? The answer has to be Mother One, but how can we explain this judgment in light of the negative consequences? Well, the virtue theorist would say that despite the bad consequences, despite her intentional risk of those bad consequences, Mother One acts responsibly, beneficially, and with some courage. Mother Two, by contrast, fails to respond appropriately to both her child and to the normative demand for responsible care which society makes on her, by basing her decision on her personal convenience. She also behaves as a free-rider, benefiting from the herd immunity maintained because other parents vaccinate their children. She is careless and callous independently of whether anyone is harmed by her choice and independently of her explicit intentions (she does not intend that her child be harmed, she is just

negligent) and therefore she is also reckless. In this case, it is fair to infer caring and carelessness/recklessness from one example because the situation has been set up as one in which there is no external pressure to act in a particular way and because the decision is the sort of routine decision that is part of the daily responsibilities of parents. As a result, these decisions can be justly seen as characteristic of the way in which the women approach risk taking on behalf of their children. It is worth noting, however, that the inference to character dispositions is only possible when we understand the process of practical reasoning that the decision makers undertake, the way in which they perceive, consider, and respond to facts of the situation, their attitudes, their values, etc. We can see from this the multi-dimensional nature of character, and this explains the comparative richness of the virtue ethicist's moral judgments when compared to those of consequentialists or deontologists.

## Habit, Education, and Moral Development

---

It seems, then, that having the right character is crucial in making the right decisions about risk. This, in turn, raises questions about the acquisition, development, and application of one's character. How does one come to have the right character and how does having the right character help with all sorts of tendencies we have toward risk?

A good place to start then is with whether a good character is a natural characteristic or one that is developed over time. At the beginning of Book II of the *Nicomachean Ethics*, Aristotle says the following:

- ▶ Virtue, then, being of two kinds, *intellectual* and *moral*, intellectual virtue in the main owes both its birth and its growth to teaching (for which reason it requires experience and time), while moral virtue comes about as a result of *habit*, whence also its name *ethike* is one that is formed by a slight variation from the word *ethos* (habit). From this it is also plain that none of the moral virtues arises in us by nature for nothing that exists by nature can form a habit contrary to its nature. For instance the stone which by nature moves downward cannot be habituated to move upwards, not even if one tries to train it by throwing it up ten thousand times; nor can fire be habituated to move downwards, nor can anything else that by nature behaves in one way be trained to behave in another. Neither by nature, then, nor contrary to nature do the virtues arise in us; rather we are adapted by nature to receive them, and made perfect by habit (NE1103a).

We are not born with good or bad or any other kind of character and it is not inevitable that we will develop one. Indeed, we do not even need one in order to produce actions (not all actions are actions from character; Butler 1988, pp 218–227). This is not to say that that we are not naturally born with certain tendencies, for example, a tendency to irascibility or a tendency to mildness of temperament, etc. Rather the claim is that natural tendencies differ from stable dispositions. Whatever our assigned lot from nature, we can become self aware of our natural tendencies, we can expose them to critical reflection, we can affirm the positive ones that guide toward virtue, and reject the negative ones that pull toward vice, and we can instil new tendencies in ourselves, until, over time, the correct ones become stable dispositions. A natural tendency toward kindness may or may not be present in particular individuals, but everyone has the option of developing the disposition to kindness, which is essential for the virtue of kindness. The tendencies with which we are born are part of what enables us to acquire

character, but they do not constitute it. Character has to be actively developed out of the raw materials with which we are born (Athanasoulis 2005).

So, moral virtue is not something we have by nature, but we are naturally possessed of the potential to become virtuous. How does the realization of that potential come about? Aristotle's answer is "through good education." Moral education, according to Aristotle, consists, in the main, in the inculcation of good habits. Often the need for habituation is taken as implying that morality is governed by the non-rational elements of the mind because it is these that need to be trained, being inaccessible to reason. However, a better way to think of the sort of habituation involved in the acquisition of good character would be to compare the process with what happens when a young person acquires a skill – playing the piano or speaking her mother tongue. The young learner learns by being required to produce simple performances of music or language and by watching others perform. In this way, she is exposed to the elements of language or piano playing and to the rules according to which these elements are usually combined (explicitly taught or inferred) and her mind sets to work analyzing these and experimenting with them. Her experiments are met with critical or approving responses by those from whom she learns as well as those affected by what she says and does. She adjusts her grasp of concepts, rules, and practices as a result. As she develops, she begins to be self-critical and takes an active role in habituating herself.

As with the acquisition of language or a skill, the acquisition of good character involves the coordination and development of a wide range of intrinsic abilities or potentialities, including reasoning, perception, emotion, desire, etc. All are capable of influencing and being influenced by each other and it simply takes good training to lay out the right tracks or set up the right relationships/patterns of interaction between them. The result is a sophisticated deliberator who can take into account relevant aspects of context and respond both intellectually and emotionally in an appropriate manner when faced with novel opportunities for choice and action. As the piano player becomes a composer and the language learner a poet, so the young agent becomes a *phronimos* – a practically wise exemplar of living well. This process of character development then involves the critical evaluation of one's natural tendencies, and the habitual work required to train one's desires to conform to the choices of one's reason, so that, when one's character is mature, the right action is chosen, chosen for its own sake, and chosen willingly and effortlessly.

Developing a good character and becoming a person of virtue is itself then a process with many built-in uncertainties, as many crucial elements in this process are subject to luck. Good moral education is crucial, but it is also rare and dependent upon contingencies of opportunity and resources. The long and difficult process of character development is vulnerable to circumstances, to the availability of good exemplars, and good influences. Furthermore, the process is as likely to be influenced by negative factors as it is by positive ones. Exposed to the wrong influences, the wrong peer group, the wrong examples, one's character may well be shaped toward vice with the same readiness that it could have, under other circumstances, been shaped toward virtue. Nussbaum points out that Aristotelian goodness of character is profoundly social and partly constituted by a capacity to interact unguardedly and generously with others (Nussbaum 1986, pp. 343ff). As a result, goodness opens the person of good character to the possibility of loss and betrayal, the experience of which is likely to be profoundly destabilizing to character. Character can be undermined even when it is fully developed and robust (Nussbaum uses the example of Euripides "Hecuba" to illustrate this claim, Nussbaum 1986, p. 397), however, like a growing plant, it is most vulnerable when it is still finding its

form. Formative experience of breaches of trust, injustice, and the loss or denial of social goods such as friendship and collective activity can all prevent character from developing well. The result may be bad or vicious character, but it might equally be weakness, confusion, and a state best described as “lack of character” (see Kupperman 1991, p. 7 for what lack of character amounts to). The process of moral development, then, is like the growth of a tender, young shoot; it will grow into a healthy, strong tree, capable of withstanding violent storms, but only if it is tended, nurtured, and exposed to the right conditions which encourage this growth. Risk and the conditions of vulnerability it creates for the development of the good character are not only acknowledged and accepted by virtue ethics, but also embraced and welcomed by it as a deep insight into the connection between vulnerability and the activity of valuing, upon which moral practice is founded.

## Decision Making

---

Thus far, we have claimed that choices to take risks ought to be morally assessed according to what those choices reveal about the character of the chooser rather than by appeal to the extent to which she respects her own and others rights or the extent to which that choice is likely to produce good consequences. We have pointed out that character is a complex multi-dimensional phenomenon, which is not only educable but also vulnerable to forces beyond the control of the agent. We have emphasized the fact that virtuous activity is activity in accordance with reason. In this section of the chapter, we will demonstrate that the role played by reason in the production of character-driven choice and action is much broader than the regulation of objective claims about what is an effective means to what. Virtue ethics recognizes that reason has a role to play in subjective valuing and emotional response. The virtue ethicist denies that personal values and feelings cannot be rationally criticized. As a result, virtue ethical analysis has the potential to penetrate aspects of risk-involving choice, upon which orthodox approaches to evaluating risk (such as those which rely upon expected utilities or stakeholder-determined priorities)<sup>5</sup> cannot even comment.

However, making decisions about risk is far from simple and unproblematic. People have natural tendencies toward risky or conservative behavior, but also toward particular, subjective interpretations of what counts as risky or conservative behavior. At the same time, reasoning about risk can become clouded with common fallacies relating to probabilities, individual variability in assessing the magnitude of risks and setting acceptability thresholds. In all this, there seems to be a conflict between an “idealized” approach to risk, involving an objective judgment based on transparent calculations of fact, and a subjective, personal interpretation of risk, which is shaped by the views, desires, and prejudices of individual risk takers. However, as we shall see, there need not be a conflict between the objective requirements of the situation and the subjective view point of the risk taker, and these two, apparently conflicting perspectives, may turn out to be far more compatible than at first thought.

Consider that in economic<sup>6</sup>, political<sup>7</sup> and even some legal/ethical (Sunstein 2002, pp. 28–53) literature about risk, which follows a broadly consequentialist approach to good risk taking, the attitude of the risk taker plays a significant but mysterious role. For example, a community that is inherently risk averse might “undervalue” a new technology (say a time machine) introduced into its environment and consequently, behave in a way that economists consider “irrational.” Economists, because they assume an essentially consequentialist theory

about risk acceptability, want to know not how something ought to be valued, but how it actually is or will be valued. To achieve this, they draw on work done in psychology concerning heuristics and how these structure thought and action. Sunstein too makes reference to psychological heuristics in his defense of expert risk decision makers. He develops the Tversky–Kaneheman heuristics to demonstrate the irrationality of risk perception and consequently of risk responses among “lay people” and uses this to argue that risk decision making ought to be left in the hands of the experts (Sunstein 2007, pp. 157–168).

So, it is widely accepted that “attitude” is one of the variables that has to be taken into account in determining the value or disvalue of outcomes, which, in turn, is necessary for deciding whether a particular risk is worthwhile. It is also widely acknowledged that “attitude” is a variable, which is difficult to track or predict. Heuristics act as correctives but, on the whole, theorists tend to simplify (reducing all subjective responses to degrees of risk aversion, thereby treating the whole of human life as if it were a gambling game) or generalize (attributing feelings like “selfishness” to all choosers and assuming that such feelings dominate in any given circumstances). When actual choosers choose differently from the ways that such theorist suggest are best, it is often claimed that these people are just bad at reasoning about risk. Because their subjective feelings and values fail to match those “arbitrarily” postulated by the theory, theorists conclude that ordinary reasoners somehow allow the irrational part of choice making to overwhelm the rational part. Subjectivity is seen as a source of error, a force that undermines reason (Lewens 2007, p. 15). Typically, those who think so take it that the subjective attitude to risk is something which is simply a psychological given which is not under much rational control and is fairly consistent across a variety of different types of risk (although it may vary according to the probability of bad results and the extent to which the consequence that is hoped for is valued). This seems to suggest a conception of risk decision making that is construes it as outside the agent’s control and open to subjective and subversive influences. Some authors think this can be overcome by removing the subjective element from risk decision making entirely and placing the responsibility for risk-involving choice in the hands of objective “experts” (Sunstein 2002, pp. 28–53). However, such claims rest upon the idea that good reasoning about risk at least approximates an exact science<sup>8</sup>, while at the same time human reasoners are at the mercy of fickle emotions and poor reasoning. This is the sort of picture of reasoning about risk that we wish to argue against.

In our view, good risk taking is the product of practically wise decision making and such decision making is not and cannot be “scientific” (for the full argument and explanation of Aristotelian “Non-Scientific Deliberation,” see Nussbaum 1986, pp. 290–317.). Aristotle argues that the right moral response is “a mean” between possible reactions. That mean is achieved as a result of sensitive perception, good practical reasoning, the well-sensitized emotional reactions, and an ability of all of these to impact upon a discerning capacity to formulate desires. The mean in each situation is different – for example, whilst rage and violent action might be thoroughly inappropriate in the case of a football fan whose team has lost, exactly the same reaction would be the appropriate response (i.e., in the mean) to someone torturing an innocent child. The variation in the Aristotelian mean has the consequence that little in the way of generalization can be made from one case to another and the correct answer will be a matter of considering the particulars of each situation. Decision making is not a matter of making or applying general, descriptive rules, rather it is intensely responsive to evaluative and normative aspects of particular situations. Specific features of the subjects involved and the circumstances in which they are involved have a crucial role to play in reasoning about what to

do and how to act well, as do features of the historical processes that brought the agent to the point of needing to make any specific risk-involving choice. It is also not “scientific” in the sense that emotional deliberation is part of the reasoning process and the emotions are compatible with and even constitutive of the agent’s reasoning. Decisions about risk that proceed from a good character involve emotional responses, which are integral to firm and stable dispositions to virtue. In what follows in this section, we will make the case for this interpretation of the Aristotelian approach and show how it enriches the explanation of moral judgments of risk taking.

Two features of Aristotelian decision making are worth discussing here: situational appreciation and practical wisdom, *phronesis*. Situational appreciation is a term coined by David Wiggins (1980) to capture the Aristotelian idea that the particular details of a case, which vary from one case to another, are crucial in ethical deliberation and it is the role of the agent to perceive these features and their ethical significance. Crucially, these particulars are situation specific, so they cannot be captured in a one-size-fits-all rule. At best, rules are rules of thumb (Nussbaum 1986; Roberts 1991; Dancy 2004), so the agent must always be sensitive to details of the particular case, which prove the exception to the rule. Situational appreciation then is the perception of the morally salient particulars of a case, which, in turn give rise to reasons to act, but these features can be difficult both to perceive and to gauge their importance relative to other features of the case. Such particulars might include the numbers of persons affected, the status and needs of stakeholders, the availability of alternatives, or features of the social context, etc, (Athanasoulis and Ross 2010). The emphasis upon attention to the particulars is in keeping with commonsense generalizations – the commonsense generalization that cases where an agent decides to expose another to risk are more problematic than cases where the agent decides to expose herself to risk is an example of different ethical significances being attributed to different cases because the “particulars of the case” (the first/third person nature of risk exposure) differ. Similarly, the reasonable claim that we have greater obligations when making any decision on behalf of a non-competent other such as a child, takes “competence” as ethically significant particular in the decision-making procedure and this seems correct. The above examples demonstrate how some sorts of particulars are relevant in a reasonably stable way, allowing the formulation of working “rules of thumb”; however, particular considerations such as “competence” and “first/third party exposure” do not occur in isolation and the Aristotelian must always allow that there are some particulars which ought to weigh differently in different situations, that “rules of thumb” are tentative and there will be contexts in which they do not apply or are even reversed leading to unconventional, but nevertheless appropriate, decisions.

A case in which context was very important in leading medical practitioners to take significant risks despite the general rule of risk avoidance was the ECMO case (Megone and Mason 2001). Clinicians were deciding to trial ECMO machines on neonates. ECMO machines provide cardiac and respiratory support to patients with severely compromised heart and lung function. In neonates, abstaining from treatment would run a very high risk of death, but the very first uses of ECMO machines also run the risk of an unconfirmed procedure, which, once initiated, could not be reversed in favor of another course of treatment. This placed clinicians deciding whether to use ECMO for the first time in a very difficult situation: not only were they making life and death decisions on behalf of incompetent others, but they had to do so under conditions of uncertainty as to the outcomes but with some evidence to suggest that the proposed treatment could avert death, their decisions were not reversible and there was no way the child could be substituted with an adult who would, at least, be able to decide for him/herself whether

the risk was worth running or not (Mason and Megone 2001). The ECMO decision is by no means clear, influenced on the one hand by the therapeutic promise to avoid a high probability of death, complicated by the lack of therapeutic alternatives once the decision is made, and constrained by the inability to substitute the child with an adult. Similar decisions may be made even more complicated because other factors of the environment turn out to be morally relevant. Similarly, they can be simplified because fewer contextual factors matter. Consider the following three slight variations of the ECMO case:

- Case 1: Medical practitioners need to decide whether to expose a pre-linguistic child to the risks of a mechanical treatment (e.g., respirator), which has not been well tested. No treatment will most likely result in death, but the treatment could result in severe brain/heart/lung damage producing severe disability. Treatment is not reversible. The case takes place in a modern hospital in London.
- Case 2: Medical practitioners need to decide whether to expose a pre-linguistic child to the risks of a chemical (drug) treatment, which has not been well tested. No treatment will most likely result in death, but the treatment could result in severe brain/heart/lung damage, producing severe disability. Treatment is not reversible. The case takes place in Somalia where many babies routinely die because there are no resources to provide them with good primary medical care.
- Case 3: Bob is deciding whether or not to put his new-born baby sister in his time machine, thereby exposing her to significant and uncertain risks.

In cases 1 and 2, it matters that the situation is bleak – the alternative to risk is death – and it matters that the relationship between the medical practitioners and the child is governed by the role and duty of doctors to their patients, which places significant weight upon saving lives. In case 3, a different kind of caring relationship is in place and thus different responsibilities arise, and there is no clear benefit to the child from participating in the experiment, in fact, there is no obvious reason why it should be a child who is exposed to the risk, an adult could have taken her place. Cases 1 and 2 are very difficult and we probably need to take further particulars into consideration before we could decide whether exposure of a child to a high degree of risk is morally required or acceptable or impermissible. Also, because the wider context of the decision and the future life of the child are also relevant considerations, it is likely that we would not arrive at the same sort of judgment in both cases. In case 3, however, things are clearer because there is an alternative that does not involve using the baby and there are reasons for Bob to be very protective of the particular child concerned. In this case, other particulars fade into the background of the case and do not play a role in decision making.

The comparison of these examples demonstrates the complexity of the particulars that can play a role in moral decision-making and the importance of becoming aware of these particulars before being able to come to a decision. The process of gradual habituation and education described in the previous section will eventually give rise to the ability to perceive morally salient features. Some features of a situation, for example, “that there is a child drowning in shallow water with no one else around to help,” are very easily perceived as morally relevant and as giving rise to obligations to act, in this case the obligation to wade into the water and pull the child out. However, many ethical situations will be far more complex than this, and agents will have to have become sensitized to perceiving and appreciating how the particulars of the case give rise to the need to act. Often, this process of sensitization will involve becoming familiar with scientific or technical aspects of a case,

which may require considerable professional expertise, but sometimes it will involve engaging one's emotional reactions and faculties of imagination. So, for example, making a decision on the time-traveling case may well involve developing a better understanding of the theoretical claims behind Bob's expectation that he can now travel through time and assessing their scientific validity at an abstract level before proceeding to a practical trial as well as the emotional attachment that he has to the project.

There is a tendency in modern choice theory to assume that self-interest is the only rational feeling and the only one whose influence upon choice deserves consideration. The Aristotelian approach to choice emphatically rejects this assumption (a contemporary argument for the rejection of this view can be found in Roeser 2010). For Aristotle, emotions are neither inimical to reason nor disruptive forces to be viewed with suspicion. The emotions can be trained and habituated to not only concur with reason but also play a role in reasoning. For example, arriving at the conclusion that one should help a homeless person find a bed on a very cold night is a process that involves both spontaneous feelings of kindness and empathic imaginings about the person's situation, which support the rational argument that help is appropriate here. Fear, sympathy, love, anger, generosity, etc., are an integral part of a proper human reaction to events. The person of practical wisdom is someone who has the appropriate emotions, to the right degree at the right time – and in doing so manifests the Aristotelian “mean.” Both habituation and reason have roles to play in ensuring that this occurs. When we decide that an emotion is “not in the mean,” then what we are detecting is a failure to balance a complex set of emotions in a way that fits the situation that one faces and recognizes the particularly morally salient features of it. Comparison, generalization, balancing, and modulating are all activities of reason assisted by emotion. This understood, it is easy to see why the subjective elements of decision making about risk are neither fundamentally irrational nor excess to requirements. In the time-machine case, it matters that when Bob makes his request of you, he is partly relying on your specific relation to him as a friend, which should lead you to view requests for assistance positively and to share in his enthusiasm about his concerns and projects. Because of this, the reasons that you have for or against exposing yourself to risk will and ought to be different to the reasons of someone who does not know him at all.

The other aspect of Aristotelian decision making we wish to highlight is *phronesis*, or practical wisdom, the state of being that underlies and underwrites all the virtues. Practical wisdom includes the ability to grasp how the different features of a situation relate to each other, which virtues relate to each case, and how the different virtues relate to each other. Practical wisdom allows the student of virtue to see not just what should be done as a matter of habit or following an example, but why it should be done, why it should be chosen, chosen knowingly, and for its own sake (Irwin 1980). This ability to reason morally allows the virtuous agent to determine the right action in a variety of different situations, as the right action will depend on the circumstances and differ according to them. The idea, that the right action will differ depending on the situation, is captured in Aristotle's doctrine of the mean. We know from the definition of virtue that Aristotle believes that the achievement of activity in accordance with “the mean” in some way requires the deployment of reason. However, it is difficult to clearly specify the precise role that reasoning plays here. One role that reasoning plays is instrumental, that is, the selection of means appropriate to ends (NE Book 7). It is now widely agreed amongst commentators that reason has *more* than an instrumental role to play in Aristotelian decision making. There has also been a great deal of

discussion about the role it plays in the selection of ends with some arguing that practical reasoning is involved in the reflective evaluation of ends (a book-length treatment of the subject can be found in Richardson 1997). John McDowell has emphasized that bringing about the integration and prioritization of morally salient particulars into decision making is very much part of the skill of the person of practical wisdom (McDowell 1998a, pp. 21–30; McDowell 1998b p. 53). Whatever the precise details, practical reason is, above all, the coordinator that enables complex psychological and active responses to match complex and demanding situations.

## To Time Travel or Not to Time Travel?

---

We began this chapter with a fictional example of risk taking, so it would seem reasonable to offer some kind of answer to that question. However, as noted earlier, Aristotle holds that we cannot expect more precision from the answer than the subject matter itself affords (NE1098a). If the subject matter is complex, diverse, and challenging, the answer has to be sensitive to these considerations and capture this complexity. If one answer does not fit all situations, it would be a mistake to demand one answer and if the answer for each situation turns out to be detailed and intricate, we should not seek to distil it to its bare essentials for the sake of simplicity. All this suggests that perhaps there is no simple, straightforward answer to whether one should participate in such an experiment. For one thing, we lack a lot of information that would be relevant in determining the salient particulars in this case. We lack information about the state of the scientific progress in terms of time travel, the details of this particular attempt, and their relative merits as compared to other scientific claims. We also lack the background information that may help us assess the participants' actions in this case. Knowing Bob will involve knowing his character and this would make it easier to judge whether, in this case, he is being reckless or cautious, appropriately ambitious or excessive, a bold ground-breaker, or a careless self-promoter. For example, you might think he is less reckless if you understand the close relationship Bob has with his sister and the huge respect he has for her decision-making capacity and technical skills despite her age. Similarly, you might judge him to be a moderate, humble, and considerate man in all other circumstances and therefore unlikely to be driven by overweening ambition. We would also need to find out about the background within which this project is taking place and whether there are any relevant checks and precautions that have been fulfilled prior to reaching this stage of experimentation. All these, as well as factors relating to your relationship with Bob and the obligations your friendship may bring to bear on this decision, are all crucially relevant details. Given the number, complexity, and depth of these details, it is not possible to give a conclusive answer to whether one should assist in this time-traveling experiment or not based on the limited information provided; however, we would like to continue making use of the case to illustrate how one might go about reasoning about it. The reasoning will be incomplete in terms of arriving at a definitive answer, but it may be helpful for illustrative purposes.

In what follows, we try to demonstrate how many of the considerations raised in this paper would go toward deliberating on whether one should risk time travel or not. We will not be able to consider all the possible different paths of reasoning that may be provoked, our aim is simply to provide a practical focus for what we have been saying about the process of virtuous choice

and action. The reasoning we consider will lay bare aspects of the reasoner's character and demonstrate ways in which that character may be flawed. It is worth emphasizing here that any moral judgment that is made is made on the basis of the whole character. Whilst we can point to flaws in the reasoning process which compromise character, there are so many possible ways in which character can be undermined that there is no possibility of reduction of the form: some immoral actions are associated with "x" flaw of reasoning and therefore to behave immorally is to include "x" in one's reasoning (or other way around, might be more straightforward but also mistaken).

The case starts with the invitation to participate in the experiment, which must produce an action-related response, for example, an intention to become involved or the opposite or no response at all. The virtuous person is likely, through a lifetime of training and development, to have the correct initial response, but for the rest of us, the struggle for virtue may mean that we have a number of different motives, which will need to be examined and endorsed or rejected as appropriate. Let us say for the moment that, in our example, you are inclined to participate in Bob's experiment. There are a variety of different reasons, which might underlie that inclination (different possible explanations of your response):

- Reason 1: Time travel might be your most cherished ambition, potentially the crown in your scientific achievements, and you simply jump at the chance to achieve it without consideration of the risks.
- Reason 2: You might be desperate to earn Bob's admiration, so keen to agree to anything he proposes
- Reason 3: You may be worried by the significant risks involved in time travel, but concerned to protect others from collaborators with less skill and understanding of the subject, so inclined to become involved so you can keep an eye on the project.
- Reason 4: You may be strongly attracted to the opportunity of helping a friend, although simultaneously unsure if the best expression of this friendship is to give Bob the assistance he asks for or to persuade him to reveal his plans to the academic community for further scrutiny before putting them into practice.
- Reason 5: You might hold the view that the risks involved are real, but worthwhile provided the time traveler is scrupulous in seeking not to alter the period to which they travel and be keen to ensure that this is the case by traveling yourself.<sup>9</sup>

Each of the above reasons for action involves the singling out of certain aspects of the available options as salient to your response, for example, the fact that it is Bob who has asked you rather than someone else, the fact that other potential participants are more vulnerable and less responsible than you, the fact that what you would be doing is time travel rather than something else equally risky but less "cutting edge," the vulnerability of "innocents" to harm as a result of your time travel, etc. In perceiving the morally salient features of a situation, you both evaluate these features and call your own responses and evaluations into question. In general, the way in which you represent the situation to yourself and the extent to which aspects of it will seem to you to be reasons for action will be the product of your educated dispositions. You are disposed to represent this complex situation to yourself in a simplified way, which emphasizes certain elements rather than others as requiring response (Butler 1988, pp. 221–227). So, in the case of Bob's request, you might see it primarily as an "opportunity for me" or as an instance of "a friend being in need of help," etc. (Butler 1988, pp. 220–221).

It is worth noting, however, that not all such representations are equally correct or equally good grounds for action. The accuracy or correctness of your own way of seeing the situation is not obvious “in the moment”, but it can be checked and changed on reflection. Reflection upon any of the above reasons for action might reveal the following sorts of “perceptual” error:

- Reasons 1 and 2: Initial considerations of personal gain, self-promotion, or scientific fervor seem to be preventing other aspects of the case such as the risks involved, the potential to harm innocents, the young and impressionable nature of Bob’s sister, etc., from being brought into the decision-making process. On reflection, one might conclude that although some concern for one’s own achievements and the regard of others are appropriate, an excess of this sentiment may cloud the judgment with respect to the risks involved.
- Reason 3: This response seems to attribute appropriate weight to the very significant risks involved and emphasizes the manipulative nature of the situation in which you find yourself. Reflection may be necessary in order to work out the extent to which one should allow oneself to be manipulated into taking risks and whether “manipulativeness” really is the most significant consideration here.
- Reason 4: This reason foregrounds the friendship between yourself and Bob, but further reflection seems to be required in order to clarify the way in which your choice is seen – in particular, your understanding of friendship, loyalty, and the requirements it places on friends are unknown, and therefore make the initial representation of the situation too vague to allow a decisive response. A second crucial feature of the situation here is the recognition of the role played by the scientific community in managing risk taking – the sort of management whose absence from Bob’s project is strikingly obvious to you, the reasoner. Recognition of this provides the reasoner with a way of making a conditional answer, which expresses enthusiasm for the project, but also takes seriously the risks involved.
- Reason 5: Here again, we have serious appraisal of the risk involved and, crucially, the recognition of the potential for managing the level of risk to which third parties are exposed and of you as better positioned than most to manage these risks. In this case, reflection might reveal that consistency requires that you be as concerned about Bob’s sister as you are about third parties and so your initial confidence in your ability to manage all the significant risks might have been overblown.

In addition to being sensitive to the particulars, reasons 1–5 above cite emotions experienced by the reasoner as a rationally relevant part of his/her response. The real prospect of time travel is likely to stimulate feelings of fear, concern for others, and excitement. In addition, you will probably have feelings that arise because of the relationship between time travel and your personal ends – you might experience joy that a long cherished and seemingly impossible scientific advance is close to realization, or you might be delighted to have the opportunity to demonstrate your loyalty and usefulness to Bob, etc. Finally, added to all of the above, will be emotions appropriate to the relationships between you (the potential risk chooser), Bob, and Bob’s sister – emotions like love, loyalty, and respect. The appropriateness of the content and intensity of particular emotions will be dependent upon your representation of the situation, your stake in the outcome, etc. We said above that emotional response

must manifest the Aristotelian “mean” and reason has a role to play in ensuring that it does. We can see from this example that being “in the mean” is not a simple matter – because it requires the balancing of different emotions as well as the modulation of individual emotions. The virtuous person will be disposed to respond with appropriately balanced and modulated emotions in a way that is similar to the disposition of the excellent language speaker to respond to a communication in the appropriate tone, with an appropriate degree of seriousness, etc. The language speaker responds in a way that is at once intuitive and underwritten by a sophisticated process of comparison and adjustment, etc., and similarly, the virtuous agent’s emotional reaction to the prospect of this time-travel adventure would be “automatic,” but nevertheless fine-tuned.

Of course, this process can go wrong – the emotions in play can get out of balance and the individual emotions may have more or less influence over the overall response than they should. Good examples of the latter are reasons 1 and 2 above. Most of us would be tempted to say that response 1 and 2 exhibit a little too much self-promotion and admiration of others, respectively, and this is cause for concern because these feelings render the agent improperly responsive to some significant risks and insufficiently careful and respectful of the needs and entitlements of others. Difficulty in balancing different emotions can be seen in reason 3 above where appropriate fear finds itself in the company of strong protective feelings toward Bob’s sister and (at least some of) the third parties who could be affected. In the scenario we are considering, protectiveness out-balances fear, but it is far from obvious that this is the right response. Similarly, in reason 5, the reasoner exhibits appropriate self-confidence but we might worry that too much weight is placed upon that feeling in the overall emotional response. Of the reasons we have been considering, reason 4 seems to be the closest to a well-balanced and appropriate emotional response blending as it does enthusiasm for the project, caution, concern for others as well as care for a friend.

Finally, it is worth noting that all of the potential reasons for participating listed above make some sort of reference to the thing, which action would be an attempt to achieve, and its purpose. The ends toward which an agent strives will range from objectives that are the products of reflection and with which she identifies, to things for which she simply has a non-rational taste. There is a very general sense in which all reasons are attempts to achieve “the best thing.” What is obvious from 1 to 5 above is that there are a lot of different conceptions of what “the best thing” is. Some have greater objective validity than others; for example, self-aggrandizement and the eternal admiration of Bob, are ends whose importance is objectively questionable. Whilst some concern for one’s own achievements is appropriate, it is hard to see why personal achievement should outweigh the value of the achievements of others that may be prevented or delayed by what you do. Similarly, the approval of others one admires is not itself an unworthy goal, but valued to an excess it will probably compromise your ability to achieve other goods that are arguably (objectively) of greater value, for example, autonomy and the well-being of others (if the person you admire is not virtuous)<sup>10</sup>. In the case of reason 3, the goal of personal safety is potentially sacrificed for the purpose of achieving a different goal – preservation of the well-being of others. Personal safety is a sensible objective and benevolent desires are noble, so here the difficulty comes in working out whether one can and should be sacrificed for the other. This is a very difficult task because good ends are not commensurable on a single scale (Nussbaum 1986; Richardson 1997). There are no infallible rules, which can help with this sort of discriminative task. Instead, skill acquired

though practice and good teaching, is necessary here. Reason 4 includes one of the ends that Aristotle thinks is constitutive of a good human life, that is, the aim to be a good friend, and combines it with something that Aristotle thinks is equally important – the goal of “good citizenship” which includes the desire for justice. Reason 5 takes as key the technological value of the project although it seeks to balance this value against the genuine disvalue of potential harm. Probably, none of the reasons we have considered give appropriate weight to all the genuinely valuable ends that ought to be in play (friendship, citizenship, technological value, the healthy and autonomous future potential of Bob’s sister, etc.); however, this examination of the sorts of ends that inform decision making and the ways in which they might be questioned is intended only to illustrate the sort of discursive process that we think a virtue ethics approach can bring to a moral debate rather than to give a complete and definitive answer.

## Conclusion

---

In this chapter, we have tried to give a brief account of the main features of virtue ethics and to illustrate how these offer a unique perspective on thinking about risk. Due to the nature of this project, the account is far from fully fleshed out and only partly defended, but hopefully it offers an insight into the theory and its contribution to making decisions about risk.

Prominent in our analysis has been the claim that the ethical life is a complex and diverse enterprise, which requires an equally detailed, flexible and situation-specific approach. One answer does not fit all, and discovering these challenging and varied answers may well be a lifelong project. This lifelong project is understood in terms of gradual, long-term, and vulnerable character development. In terms of risk, this means that we should shift the focus from the consequences of one-off, and often extreme, cases, to a broader view. We should consider the nature of the decision to take a risk, the type of character this decision reveals, and the life within which this character is displayed. The evaluation of the decision to risk must reflect the morally salient particulars of the situation and how these are ordered by the person who exhibits *phronesis*. If this approach results in a less prescriptive and less direct answer, this is no cause for complaint. For, it is the nature of ethical judgments that they are difficult to arrive at and require thought and effort. We shouldn’t expect anything less of ethical judgments involving risk.

## Further Research

---

In this paper, we have examined the essential role of character in good decision making about risk. In a previous paper, we have considered the way in which aspects of context influence decision-making about risk (Athanassoulis and Ross 2010). These two papers set up a framework for a virtue approach to risk taking. There is much work to be done to fill out this framework. In particular, we have emphasized that a virtue approach would need to consider particular issues on a case-by-case basis and there is work to be done looking at individual risk problems in particular disciplines. For example, questions about the circumstances in which it is acceptable to expose research subjects to risk or questions concerning what sort of risk taking in business is

ethical, etc. In addition, this paper has emphasized the role of education for virtue and more work is needed to fill out an account of what it takes to educate virtuous risk takers and to look at how such education can be incorporated into the professions.

## Notes

---

<sup>1</sup>We have chosen time travel as our example of risk decision making because it has the advantage of being unfamiliar and challenging to readers from all disciplines and yet embodies many of the dilemmas that will arise when making decisions concerning the use of any risky technology.

<sup>2</sup>Deontologists differ over what defines moral rights and duties. Some think they are determined by divine command (Quinn 1999, p. 53), and others by political consensus (Hobbes 1660, Chap. 13; Rawls 1971, Book I). The most influential of the deontological theories has been Kant's – he argues that we can work out what our rights and duties are by appeal to "the categorical imperative" which is the logical consequence of seeing ourselves as valuable because we are agents and, as a matter of consistency, seeing all other agents as equally valuable (Kant 1785).

<sup>3</sup>Of course, not all situations involving risk are instigated by an agent, for example, many are the result of natural forces over which we exercise no control, but in this chapter, we are interested in moral judgments and responsibility for one's decisions so we have limited 'the discussion to these kinds of cases of risk. For more on this, see Athanassoulis and Ross 2010.

<sup>4</sup>This example is adapted from Hansson 2003, p. 295. The reason for the changes is that the individual in Hansson's original example is a repairman with the skills to effect the repair. It seems to us that selecting this person in particular raises questions about one's obligation to expose oneself to danger when the risk of doing so is part of one's employment obligations, which were freely entered upon. Such considerations, as well as more general questions regarding the selection of the individual who runs the risk, for example, whether he/she is a volunteer, whether he/she has any role obligations in this respect, etc., make a difference to the moral evaluation of the decision.

<sup>5</sup>In these views, "the passions" which include both desire and emotion are the product of causal processes that are little or are of no interest to philosophers. They take it that these elements can best be catered to by studying the patterns of reactions that people actually have and using these as a measure of value.

<sup>6</sup><http://homepage.newschool.edu/het/essays/uncert/aversion.htm>; <http://hadm.sph.sc.edu/courses/econ/Risk/Risk.html>; <http://moneyterms.co.uk/risk-aversion/>

<sup>7</sup>Rawls, for example, feels the need to stipulate a level of risk aversion, which he considers to be reasonable in arguing for the MAXIMIN strategy (Rawls 1971, pp. 123–133)

<sup>8</sup>The basic idea is the consequentialist one that reasoning about risk is a matter of marrying probabilities and value/disvalue of potential outcomes and then selecting the combination that comes out with the best score.

<sup>9</sup> This list is not intended to be exhaustive, merely illustrative.

<sup>10</sup>Aristotle thinks that what is valuable for its own sake is an objective matter – that the good life for a rational being has a substantive nature, which requires the pursuit of particular ends – ends such as friendship, political participation, justice, contemplation, creative achievement,

etc. So, reason enters into “end-adoption” as a criterion –  $x$  is only a good end if its pursuit in some way contributes to our living the lives of rational beings; if it is an end whose pursuit will exercise and develop our rational (as opposed to arational) natures.

## References

---

- Aristotle (1925) *The Nicomachean ethics* (trans: Ross D). Oxford University Press, Oxford
- Athanassoulis N (2000) A response to Harman: virtue ethics and character traits. *Proc Aristot Soc* 100(2):215–221
- Athanassoulis N (2005) *Morality, moral luck and responsibility*. Palgrave, Basingstoke
- Athanassoulis N, Ross A (2010) A virtue ethical account of making decisions about risk. *J Risk Res* 13(2):217–230
- Butler D (1988) Character traits in explanation. *Philos Phenomenol Res* 49(2):215–238
- Dancy J (2004) *Ethics without principles*. Oxford University Press, Oxford
- Hansson SO (1993) The false promises of risk analysis. *Ratio* 6:16–26
- Hansson SO (2003) Ethical criteria of risk acceptance. *Erkenntnis* 59(3):291–309
- Hobbes T (1660) *The Leviathan*. <http://oregonstate.edu/instruct/phl302/texts/hobbes/leviathan-c.html> – Chapter XIV. Accessed July 2010
- Hursthouse R (1999) *On virtue ethics*. Oxford University Press, Oxford
- Irwin TH (1980) Reason and responsibility in aristotle. In: Rorty AO (ed) *Essays on Aristotle's ethics*. University of California Press, Berkeley, pp 117–156
- Kant I (1785) *Groundwork of the metaphysics of morals* (trans. Gregor MJ, 1998). Cambridge University Press, Cambridge
- Kupperman J (1991) *Character*. Oxford University Press, Oxford
- Lewens T (ed) (2007) *Risk: philosophical perspectives*. Routledge, Abingdon
- Mason S, Megone D (eds) (2001) *European neonatal research*. Ashgate, Aldershot
- McDowell J (1998a) Some issues in Aristotle's moral psychology. In: *Mind, value and reality*, Harvard University Press, Cambridge MA, pp 23–49
- McDowell J (1998b) Virtue and reason. In: *Mind, value and reality*. Harvard University Press, Cambridge, MA, pp 50–73
- Nussbaum M (1986) *The fragility of goodness: luck and ethics in Greek tragedy and philosophy*. Cambridge University Press, Cambridge/New York
- Quinn P (1999) Divine command theory. In: LaFollette H (ed) *The Blackwell guide to ethical theory*. Blackwell, Oxford, pp 53–73
- Rawls J (1971) *A theory of justice*. Harvard University Press, Cambridge, MA
- Richardson H (1997) *Practical reasoning about final ends*. Cambridge University Press, Cambridge
- Roberts RC (1991) Virtues and rules. *Philos Phenomenol Res* 51(2):325–343
- Roeser S (2010) Intuitions, emotions and gut feelings in decisions about risks: towards a different interpretation of “neuroethics”. *J Risk Res* 13:175–190
- Sunstein C (2002) *Risk and reason: safety, law and the environment*. Cambridge University Press, Cambridge
- Sunstein C (2007) Moral Heuristics and risk. In: Lewens T (ed) *Risk: philosophical perspectives*. Routledge, London, pp 156–171
- Wiggins D (1980) Weakness of will, commensurability, & the objects of deliberation & desire. In: Rorty AO (ed) *Essays on Aristotle's ethics* (1980). University of California Press, Berkeley, pp 241–266

# 34 Risk and Trust

Philip J. Nickel · Krist Vaesen

Eindhoven University of Technology, Eindhoven, The Netherlands

<b>Introduction</b> .....	<b>858</b>
<b>History</b> .....	<b>858</b>
<b>Current Research</b> .....	<b>859</b>
Philosophical Conceptions of Trust and Risk .....	859
Two Relationships Between Trust and Risk .....	859
Trust as Interpersonal Staking .....	863
Trust Beyond Rational Calculation .....	864
Trust as a Moral Attitude .....	866
Scientific Approaches to Trust and Risk .....	868
The Science of Trust as Rational Risk Taking .....	869
The Science of Trust as Non-Calculative .....	871
The Science of Trust as a Moral Attitude .....	872
<b>Further Research</b> .....	<b>873</b>

**Abstract:** Philosophical conceptions of the relationship between risk and trust may be divided into three main families. The first conception, taking its cue from Hobbes, sees trust as a kind of risk assessment involving the expected behavior of another person, for the sake of achieving the likely benefits of cooperation. The second conception of trust sees it as an alternative to calculative risk assessment, in which instead of calculating the risks of relying on another person, one willingly relies on them for other reasons, e.g., habitual, social, or moral reasons. The third conception sees trust as a morally loaded attitude, in which one has a moral expectation that one takes it to be the responsibility of the trusted person to fulfill. In the context of interpersonal relationships this attributed moral responsibility creates spheres perceived to be free of interpersonal risk, in which one can pursue cooperative aims. In this chapter, we examine how these three views account for two *prima facie* relationships between risk and trust, and we look at some empirical research on risk and trust that employs these different conceptions of what trust is. We then suggest some future areas of philosophical research on the relationship between trust and risk.

## Introduction

---

This chapter describes some prominent philosophical connections between trust and risk. Section **● History** briefly describes historical thought about these connections. In section **● Current Research**, we describe current philosophical research on trust, including relevant empirical work on trust, attempting to understand how trust helps people pursue welfare cooperatively under conditions of risk. We conclude by indicating, in the final section, areas of future philosophical research on trust and risk.

## History

---

There are only a few significant references to trust in ancient Western philosophical sources, such as a work by the “anonymous Iamblichus” (Gagarin and Woodruff 1995, cited in Hardin 1998). Trust thought of as a practical solution to situations of risk has its modern theoretical origins in Hobbes. Humans outside of civil society, Hobbes argues, are vulnerable to risks both from the harsh natural environment and to risks from other people. It is rational to protect one’s own life, Hobbes argued, but for this very reason rational cooperation is not generally possible, because in many situations it will be rational to exploit others’ willingness to cooperate, and to expect similar exploitation from others. Therefore, outside of social and political structures it is rational to treat all other persons in a warlike way, cooperating with them only if it is possible to guarantee that they will not do harm to oneself. Many cooperative goods will then be impossible to obtain (Hobbes 1968 [1651], p. I.13). For this reason, Hobbes thinks it is practically necessary to establish a powerful political authority that will coercively enforce the terms of promises and ensure the possibility of mutual cooperation (Hobbes 1968, p. I.14). Hobbes calls *trust* the attitude of rational confidence in the reciprocal cooperation of others that is safeguarded by such a powerful authority (Hobbes 1968, p. I.15).

The particular way that trust figures in Hobbes’ political philosophy was not addressed explicitly by his many contemporary detractors (they focused more on his conception of the rational person and the natural law). Indeed, the notion of trust was relatively unimportant in moral, political, and social philosophy throughout the modern period, except for a minor role in

the political philosophy of John Locke (Laslett 1988; Dunn 1984). When trust reemerged as a philosophical topic in the twentieth century, it did so in three main areas. Some scholars picked up where Hobbes left off, seeking to understand how cooperative relations going against immediate self-interest but beneficial for social well-being are strategically and motivationally enabled under conditions of risk (Held 1968; Coleman 1990). A second group sought a more wide-ranging understanding of trust as an explanatory concept in social theory and social science, thinking of trust as an emergent property of societies in modernity (Luhmann 1979). Finally, a third group, beginning with the work of Annette Baier, sought a moral-philosophical account of trust. Baier criticizes the other traditions for failing to give a satisfactory normative ethical account of trust outside of contractual or quasi-contractual relations. She points out that trust relationships are characteristic, perhaps especially so, of relationships that cannot easily be thought of as contractual or exchange based, e.g., parent–child relationships and other relationships among intimates, and that here and elsewhere the question of whether trust is good or bad is not merely strategic (Baier 1986). Since all three of these conceptions of trust are still actively relevant to the relationship between trust and risk, they will be discussed in the next section on current research.

## Current Research

---

### Philosophical Conceptions of Trust and Risk

---

In the first subsection below, we give some initial definitions of risk and trust, and then set out two apparent conceptual relationships between risk and trust. This will provide an impression of the relationships to risk that should be explained by a philosophical theory of trust. We then proceed to discuss how three families of philosophical conceptions of trust explain and interpret these relationships.

### Two Relationships Between Trust and Risk

Compared to the concepts of risk and risk perception, the concept of trust is relatively problematic and unsettled. Following Hansson, we define risk as a possible but not certain future harm, or the probability of such a harm, or the expected disutility of such a harm (Hansson 2004). Risk perception is the mental representation of a risk, as realized in emotions such as fear and in cognitive states such as prediction.

It is sometimes useful to draw a distinction between risk and epistemic uncertainty. On this understanding, risk is one kind of uncertainty: a quantified or well-characterized uncertainty. Other kinds of uncertainty are less well quantified or characterized. For example, in a poker game, the likelihood that a given poker hand will lose is a well-characterized possibility, given a fair game. It can be assigned a numerical value. The likelihood that the hand will lose because the game is not fair, however, might not be easy to characterize or quantify. We sometimes want to distinguish these two factors by calling the former a risk and the latter an uncertainty. In risk perception, these two factors are sometimes difficult to separate.

Trust is a concept used variously by research in psychology, sociology, management studies, philosophy, economics, and political science. There are many conflicting accounts of what it is, how it arises, and what kind of explanatory work it might perform in an account of

interpersonal relations. Even within these fields there is disagreement about how to define and investigate trust (McLeod 2002; Hardin 2006). Because of these disagreements, we will settle on as neutral an initial definition of trust as possible. Trust is *a disposition willingly to rely on another person or entity to perform actions that benefit or protect oneself or one's interests in a given domain*. This definition departs from some prevalent accounts of trust in several ways. First, it leaves out the distinctively moral character that the attitude of trust is often conceived to have by philosophers. (This view will be discussed in the section [Trust as a Moral Attitude](#)). Second, it fixes two objects of trust: the agent and the actions-within-a-domain that are to be performed by the agent. This notion of trust, containing a trustor, a trustee, and a domain or an action, is referred to in the literature as “three-place trust” (Baier 1986) and is already implicit in one of the earliest contemporary philosophical treatments of trust (Horsburgh 1961). The three-place account is more concrete than accounts of trust that view the object of trust as a measure of general confidence in the good intentions of other persons or institutions. The third point relates to the *willingness* of trust. When one's disposition to rely on another is only due to coercion or a lack of alternatives, it is not a trusting disposition. In cases where one trusts, one has some independent reason for doing so, relating to the qualities of the trusted entity.

The definition, however, leaves open what characteristic reasons support trust. Most accounts of trust specify this in some way or other, e.g., by saying that it is characteristically based on social information, such as group membership, relationship, reputation, and power. This point about the reasons for trust is important for our understanding of the relationship of trust and risk. It is important for understanding the first of two main conceptual relations we use to see how theories of trust address the issue of risk. Suppose we are thinking of hiring WDC Dredging Company to conduct a small dredging project. The *Handbook of Dredging Engineering* states that “Characterizing the environmental risk posed by dredged material requires information on the likelihood that organisms will be exposed to contaminants and the probability that such exposure will produce adverse effects in the environmental receptors (organisms) of concern. The uncertainty associated with these estimated risks has been high” (McNair 2000, p. 22.39). Special techniques have been developed to reduce these risks under different conditions. In order rationally to rely on WDC to carry out the project, however, we need not quantify these risks and uncertainties ourselves, reaching a conclusion about how they should be addressed. We could instead confidently count on WDC to know and apply appropriate methods, and to abide by relevant laws and industry practices – in a word, we could trust them. Our reasons for this would have to do with WDC's reputation, with the existence of regulatory bodies, and with the knowledge that WDC wishes to protect its future reputation. They would not normally depend on specific information about probable dangers due to this particular dredging job.

This suggests that the question whether it is a good or bad idea to rely on another person or entity to perform a given action can be answered in two ways. One way is by evaluating the risks associated with that action. The other is by determining whether the other person or entity is trustworthy. For example, whether it makes sense to rely on WDC to dredge a harbor safely and without environmental damage partly depends on the quantifiable risks that are associated with this agent performing this activity. But, it also depends on the trustworthiness of WDC relative to other companies. Therefore, in a situation like this, risk evaluation and determination of trust are in some sense *different answers to the same question*: whether to rely on the company to perform a given action. Whereas risk evaluation focuses on the probabilities

associated with the underlying activity, determination of trust focuses on the qualities of the entity (in this case, the company) such as its competence and motives, as influenced by its social and normative standing. In case one of the answers is incomplete or inadequate, it may be possible to answer the question in the other way: where risks cannot be easily quantified under the circumstances, one can ask whether the company is trustworthy; where the company is unknown, one can attempt to quantify the risks independently. This is echoed in a recent survey article by a group of psychologists: “There are two general ways [that individuals predict cooperation in others]. One, *trust*, is defined as the willingness, in expectation of beneficial outcomes, to make oneself vulnerable to another based on a judgement of similarity of intentions or values; the other, *confidence*, is defined as the belief, based on experience or evidence, that certain future events will occur as expected” (Earle et al. 2007, p. 30; see also Luhmann 1979; Seligman 1997; and Deutsch 1977 for other versions of this distinction).

A corollary is that actively controlling risks tends to reduce trust rather than increasing it. To the extent that we attempt to quantify and control the risks and uncertainties of this dredging job and choose the proper methods ourselves, this reduces the scope of our trust in WDC by reducing the extent of our reliance and vulnerability, even though the action that we essentially rely on them to do – dredging the harbor – remains the same. If we attempt to reduce our risk by taking out insurance, this reduces the extent to which our well-being depends on WDC’s fully adequate performance. Or suppose we want to be as sure as possible that WDC will perform safely, and to this end we do a detailed investigation into this company’s entire history of activity and the safety of the methods they use, and moreover formulate a plan to monitor the stages of their dredging operation. This would reduce our epistemic reliance on WDC. In such cases, where we take active steps to control or reduce our risks, epistemically and practically, it might be appropriate to remark that we *do not fully trust* WDC, or at any rate that after going through these steps we do not need to trust them as much. It seems, then, that as our own knowledge and control of the risks associated with reliance on another entity increase, our independence increases, our reliance and vulnerability decrease, and our (need for) trust in the other entity decreases. Hence in this way, too, the characteristic reasons for risk taking and for trusting are different from one another. This, then, is the first main conceptual point about risk to be explained by theories of trust: the characteristic reasons for risk evaluation and for trust are different, and to some extent mutually exclusive.

The second main apparent conceptual relationship to be explained by a theory of trust is that high background risk makes trust more salient, but also more difficult to justify. Suppose Q is the organizer of a protest group under a repressive government. Q faces the risk that if the group is discovered, she will go to prison. In order to expand membership in the group and coordinate group actions, she must tell others about the group and count on them to carry out activities on the group’s behalf in strict confidence. In such a situation, Q must rely on other people to protect her interests and those of the group. Her willingness to do this goes along with her trust. It seems that in cases like this where risks are high, but where reliance on others is essential to a valued activity, well-grounded trust is of great importance (Boon and Holmes 1991, cited in Das and Teng 2004, p. 87). Trust often coexists with risk and is made more salient by it, because the need for trust increases as the risks associated with a valuable act of cooperation increase. On the other hand, in trusting one wants to minimize risk as much as possible by finding the most competent, cooperative people available and the most conducive background conditions. In our example, Q would be rational to trust those individuals least likely to compromise her interests while helping fulfill her goals and those of the group.

It might be rational, e.g., to collaborate with family members or close associates of people already in the group, because these persons have more to lose if they defect. Because the stakes are high for Q, she must be very careful about whom she trusts and under what circumstances. Thus, the risks associated with a cooperative act make trust more important, but also more difficult to adopt on careful, stringent grounds.

One way of justifying this point is to draw on intuitions from recent epistemology about the relation between knowledge ascription and practical stakes. The intuition is that the greater the practical stakes that hang on the truth of some claim P, the more difficult it is to have knowledge or sufficient justification for believing P. Here is a version of a standard example used to motivate this intuition:

- ▶ It's late on Friday afternoon, and raining hard. Lo and her next-door neighbor Hi are [separately] thinking about going out to the bank, but wondering whether the trip could be postponed until tomorrow. Both of them can remember a recent Saturday visit to the bank, but neither of them has any further information relevant to the question of whether the bank will be open tomorrow. Nothing much is at stake for Lo—her banking errand could be done any time in the next week—but for Hi the question has burning practical importance. Hi knows he must deposit his paycheck before Monday, or he will default on his mortgage and lose his home (Nagel 2008, p. 279; the original example is found in DeRose 1992).

Among epistemologists, a widely adduced intuition about this case is that whereas Lo knows (or is justified in believing) that the bank is open on Saturday, Hi does not know this, even though Hi and Lo have the same evidence. The only difference between Hi and Lo is what they have at stake concerning the question whether the bank is open, hence this is taken as the key difference that explains why one of them has sufficient justification for knowledge and the other not. High-stakes practical reasons raise the bar of sufficient justification. Assuming we can extrapolate from this type of high-stakes case to other cases, such as Q's question whether a given person will keep secrets about the protest group from the authorities, it appears that sufficient justification for trust is difficult to come by partly because the epistemic standards for knowledge or justified belief are higher in Q's high-stakes situation, although the data supporting this intuition and the interpretation of the alleged intuition are disputed (Buckwalter 2010; Nagel 2008).

There is also the more mundane principle that where more is at stake, there is often more reason for any given individual relied upon not to perform as expected, and consequently more reason not to trust them. People in Q's country take a much greater personal risk by participating in a protest group than those in countries without an oppressive government, so they need to have special motives or qualities such as courage in order to perform as Q expects of them. What is needed for trustworthiness in this high-stakes context may therefore be more demanding. Since the characteristics underlying trustworthiness are more rare, it will be more difficult for Q to justify trust. It is also distinctive of this example that those who are trustworthy for Q may have special reason to conceal the very characteristics that would indicate this to others. There also may be government agents who pretend to have such characteristics in order to infiltrate and undermine the protest group. Thus Q's epistemic task is difficult indeed.

These, then, are two main apparent conceptual relationships between risk and trust that one hopes a theory of trust will elucidate: on the one hand, extensive risk evaluation makes trust less relevant; on the other hand, when the risks are greater, trust is more difficult to justify. In the remainder of this section, we will set out three prevalent approaches to the theory of trust and see how well they explain and justify these apparent relationships.

## Trust as Interpersonal Staking

The first approach under consideration sees trust as a special instance of rational risk taking. As Sztompka writes, “Trust is a bet about the future contingent actions of others” (1999, p. 25; see also Gambetta 1988). Suppose Q considers inviting Ron to join her protest group. If Ron is trustworthy, his membership in the group will bring a significant added benefit to Q. If Ron is not trustworthy, inviting him could be very detrimental to Q. Thus, in considering whether to invite Ron, Q must weigh the likely benefits of inviting him against the possible risk that if she does so, Ron will betray her and the group. This can be regarded as a calculated choice, or a kind of bet Q is in a position to make on Ron’s future behavior. Some have regarded the willingness to make such an interpersonal bet as the essence of trust. There are important methodological reasons for considering trust in this way, such as its ability to help predict and explain macro-level social effects as a function of the cooperative behavior of individuals with limited options and interests, where the values of different options for individuals (including the option of reliance on others) can be thought of as preference rankings or subjective expected utilities (Coleman 1990, Ch. 1).

This type of view does an excellent job explaining the second of the conceptual points set out above, why the difficulty of justifying trust increases with the risks. This is due to the obvious fact that on such a view one’s dispositions to trust are based on expected utilities, which take into account both the likelihood of performance, and the stakes to be gained through performance and lost through nonperformance. The increasing benefits of performance can increase the willing disposition to rely on another; similarly, the increasing harms of nonperformance can decrease this willing disposition. This means that the willing disposition is not exclusively dependent on the *belief that another person will perform* (is likely to perform). It is just as dependent on the *stakes* that are attached to performance and nonperformance, where the expected utility of reliance given these stakes is compared with a situation in which one relies on somebody else, or does not rely on anybody. Such a view is therefore quite different from a doxastic (belief based) view of trust according to which the willing disposition is based on a *belief* that the entity relied upon will perform in a certain way. One need only believe it is likely enough that the entity will perform to make it worthwhile, given the utilities of performance and nonperformance, to rely on him (Nickel 2009). In cases where there is little to lose given nonperformance, one need not believe strongly that a person will perform in order to feel comfortable relying on him. But that is not Q’s situation. She has a lot to lose through Ron’s nonperformance, so before she can willingly rely on Ron she needs information that more fully ensures that Ron will perform, and this increases the justificatory burden for a belief in his trustworthiness.

However, this view does not do much to explain the first of the conceptual points set out above, the mutual exclusivity between reasons for trusting and detailed risk assessment. To emphasize this, consider a thought experiment involving the dredging job again. Suppose I have conducted detailed research on the history of all dredging jobs in my region, including those of WDC. I also come to learn almost perfectly the scientific appraisal of the methods used by WDC. I come to have as much knowledge about these topics as anybody ever has before. Suppose, too, that I have taken out a large insurance policy to protect me in case WDC does something wrong for which I am liable. I also form a detailed plan for monitoring WDC. Now on the “staking” view of trust that we have been considering, this gives me perfect grounds for the attitude of trust: the more I can guarantee performance and safeguard against

nonperformance, the more purely I can trust. But this is exactly the opposite of the first intuitive link discussed above, according to which these are not the characteristic reasons for trust.

Some advocates of this type of approach have specified that trust relates only to reliance undertaken on the basis of certain sorts of reasons. One prominent view in particular adopts the condition that Q's trust is characteristically based on Q's belief that the one to be relied upon (Ron, in this case) cares about Q's interests as part of Ron's own interests: Q believes that her interests are *encapsulated* in Ron's (Hardin 2006). This moves some way toward explaining the distinctively interpersonal nature of reasons for trust. According to this encapsulated interest view of trust, possible reasons for thinking one's own interests are encapsulated in another person's interests include the thought that it is for their own strategic future benefit to perform as expected, that they are subject to sanctions if they do not perform as expected, or that they are fair and law abiding. The risks involved in trusting are interpersonal risks: they therefore depend on, among other things, the motives of the trusted person.

However, Hardin does not hold that efforts to quantify risks objectively or control them result in a reduction of trust. Indeed, on his view they would seem to enhance trust, since trust is for him still calculative. Reacting against this view, the sociologist Eric Uslaner distinguishes between strategic trust and moralistic trust (1992). Strategic trust is based on rational assessment and verification; in moralistic trust, one relies on others without detailed risk assessment and without expecting anything else in return, as in Uslaner's example of a fruit stand owner who leaves the stand unoccupied, expecting others (mostly people he does not know and will never meet) to pay on the honor system for what they take (Eric Uslaner 1992, pp. 14–15). There are many situations in which we seem to trust unknown others by default, without frequent disappointment (Verbeek 2002, pp. 140–141). Decisions to trust have been shown not to be based on risk information in some studies (Eckel and Wilson 2004). Indeed, some have even argued that we are under some kind of weak moral obligation to regard others in this way, until they are proven unreliable (see Thomas 1978). The question then is what, if anything, could make it epistemically rational to adopt a trust attitude on these grounds. So as not to lose a connection with a quantitative behavioral theory, Das and Teng suggest that this appearance of non-calculative, default trust is an illusion: estimates of the likelihood of a trusted person's performance can be posited below the surface as "unconscious calculations" about the good-will and competence of others, based on estimates of risk (2004, p. 99).

## Trust Beyond Rational Calculation

The second approach sees trust as an interpersonal belief that is essentially nonrational or non-calculative, involving to some extent a leap of faith. Such a view can originally be traced to Luhmann (1979), who held that trust is primarily a means of reducing complexity by relying on others instead of making an independent assessment of risk. As Guido Möllering sets out the idea, many philosophers of a rational choice orientation are led a certain distance in this direction by the need to acknowledge the occasional therapeutic or future oriented, rather than evidential, nature of trust. Placing trust is sometimes used as a means of cultivating reciprocal engagement and cooperation; in many cases it can actually help produce the effect of performance, even though there was no specific prior evidence this would occur. In this way trust is

“reflexive” (Möllerling 2006, pp. 80–84). But according to Möllerling, this does not go far enough, because this still assimilates trust to a calculation based on past performance and future consequences, now taken in a slightly broader sense. Möllerling advocates a more radical account that instead views trust as more fundamentally nonrational, according to which it is truly a “leap of faith” in which one suspends calculations about benefits and risks to some degree. On this view, trust does not just reduplicate an explanation that could already be provided by the rational choice paradigm, but can instead be usefully applied to empirical phenomena of cooperation that do not easily fit that paradigm, such as patients’ suspension of rational considerations when electing a surgery (Möllerling 2006, pp. 188–189). Another expression of this idea is that trust involves *optimism* about the performance of others (Jones 1996), an attitude in which despite the presence of risk a good outcome will be achieved (Nooitboom 2002).

If we are reluctant to postulate such a motivation for cooperation beyond a subjectively rational basis, then we can explicate the view more prosaically by claiming that the essential reasons of trust are substantially interpersonal, social, and/or moral instead of calculative. It is important not to conflate this non-calculative view of trust with the view that trust is emotive. On the one hand some emotions, e.g., fear, can be regarded as affective “calculations” of risk, and on the other hand some socially based judgments have none of the affect characteristic of emotional states. Thus, trust is properly based on the following kinds of considerations:

- The desire of the trusted to engage in future interaction with the trustor
- The desire of the trusted to protect their future reputation (Pettit 1995)
- The security of the institutional context in which the trusted person operates
- The moral qualities of the trusted person (see section ➤ [Trust as a Moral Attitude](#) below)
- General confidence that things will work out no matter what happens
- The ethical desirability of trust itself

That some of these are reasons for trust is already acknowledged by advocates of the staking model. But their acknowledgement is subject to the understanding that such elements can be converted into calculative factors in one’s judgment of the benefit and likelihood of different outcomes. Taking the reasons in this way disregards one of the principal motivations of the view under discussion, which is that trust is “a functional *alternative to rational prediction* for the reduction of complexity. Indeed, trust is to live as if certain rationally possible futures will not occur” (Lewis and Weigert 1985, p. 976, italics added; cited in Möllerling 2006, p. 83). Certain risks are therefore excluded from consideration at the moment of deciding to trust. On such a conception perfectly central, normal cases of trust would include situations in which no good calculative basis for trust is available, nor is employed (in terms of a concrete track record or an estimation of general reliability in similar situations), but willing reliance on another nonetheless occurs.

Such a view does a good job explaining the first of our earlier intuitive claims about the relationship between trust and risk, that trust formation and risk evaluation seem like quite different processes, and that risk evaluation and control tend to make trust irrelevant or out of place, rather than increasing it. Since trust is non-calculative, risk evaluation and control tend to push it aside. As for the second of our earlier claims about risk and trust, the non-calculative view of trust takes a paradoxical stance. On the one hand, it grants the normative claim that from a restricted rational choice point of view, justifying trust becomes more difficult as the

stakes increase. But it holds that trust can proceed without this type of justification, either on the basis of sheer optimism, or on the basis of interpersonal, social, or moral reasons. In order to have an interesting field of application in the real world, such a conception of trust invites the corresponding empirical hypothesis that the occurrence of willing reliance on others is not directly inversely proportional to the magnitude of risks perceived to be associated with this reliance. Evidence on this point is ambivalent. Empirical studies of the prisoner's dilemma and other similar game situations seem to indicate that people sometimes willingly rely on others even in the absence of calculative reasons to do so (Möllering 2006, p. 38; Hardin 2006, pp. 50–51). This suggests that such a non-calculative conception of trust may well have an interesting domain of application.

However, there are still philosophical worries about such a view. For example, Pamela Hieronymi has argued that the attitude of trust is conceptually constrained by the fact that it inherently answers the question “Is the other person trustworthy?” (Hieronymi 2008). In much the same way as belief in P answers the question “Is it the case that P?” and an intention to φ answers the question “Is φ to be done?” and the attitude of trust toward some person R answers the question “Is R trustworthy?” Although Hieronymi leaves open what kinds of considerations close the question of whether R is trustworthy, her account implies that in really adopting the attitude of trust, one is always committed to having sufficient reason for thinking that the other person is trustworthy. To the extent that her argument is plausible, it tethers the notion of trust to the reasons for thinking another person trustworthy. Genuine trust is not compatible with the presence of severe, realistic risks associated with reliance that are known to the truster and have not been ruled out, controlled, or hedged.

## Trust as a Moral Attitude

The third approach to the concept of trust sees it primarily as a moral attitude among intimates and community members, defining social spheres within which cooperative actions are safely pursued. A good example of this approach can be seen in the work of Caroline McLeod, who thinks that any notion of trust should first accommodate “prototypes” or central examples such as “child-parent relationships, intimate . . . relationships, and professional-client relationships” (2002, p. 16), and then address theoretically useful but more peripheral examples such as self-trust and trust between strangers. According to McLeod, the prototypical examples of trust are best seen as involving an ascription of certain moral and relationship-regarding qualities to the trusted person, such as moral integrity, and a shared perception of the relationship itself. These qualities ensure that the person will have *goodwill* toward us, a characteristic belief of trust first emphasized by Annette Baier (1986). The focus on trust as connected with the quality of intimate relationships can be traced back to the developmental psychologist Erik Erikson (1950), who held that the quality of the mother–child bond determines trust dispositions that are crucial for normal social development.

Baier holds that trust essentially involves remaining vulnerable to the actions of another person, partly due to the discretion given to that person within a domain of activity (Baier 1986). This vulnerability is especially present in intimate relationships for several reasons: because the domain of shared activity is large; because the discretion given is considerable; and because the relationship itself is often valued intrinsically, so that something extra is at stake should a failure or breakdown occur. This is sometimes manifested in a disposition to *feelings* of vulnerability

as well, something noted by empirical studies of partner trust (Kramer and Carnevale 2001). On Baier's view, we willingly increase our vulnerability to another because we are confident of the *competence* and the *goodwill* of the other person, although we often do not consider these factors consciously. Vulnerability provides a link with risk on Baier's account. Correspondingly, feelings of vulnerability or security provide a link with risk perception.

In addition, the *state* (as opposed to the feeling) of vulnerability is conceptually related to risk. In talking about the causal components of risk, it is common in technical literature on risk management to distinguish between vulnerability, exposure, and hazard (Renn 2008, p. 69). Vulnerability is a (relational) property of a thing that makes it potentially susceptible to damage or harm, should exposure to a hazard source occur. Baier does not draw this distinction between vulnerability and exposure, but it is instructive to do so. We might even restate Baier's view by saying that trust involves, not increasing one's inherent vulnerability, but instead increasing one's *exposure* to a potential hazard source – in this case another person – by giving them discretion over something important. This, in turn, creates a feeling of vulnerability.

The reason philosophers often claim that the attitude of trust has an essential moral dimension is that trust seems intuitively to commit one to certain characteristic morally tinged responses in cases where another person fails to perform as they were trusted to do. Moral philosophers are careful to distinguish two fundamentally different kinds of expectation that we can take toward other people when we rely on them (Holton 1994; Faulkner 2007). *Predictive* expectations are simple beliefs that the future will go one way or another. They do not inherently involve a personal or emotional commitment to its going one way or another. When one relies on the future being a certain way in this predictive sense, one is at most disappointed if one's expectation is not met. *Normative* expectations, on the other hand, involve a prescriptive judgment or assumption that the future *should* go a particular way. If the future does not go this way – if a person does not do what she has agreed to do, for example – then moral disapprobation and other richly evaluative or moral judgments are appropriate. In cases where one willingly relies on another person, but is not disposed toward any moral judgment such as blame or betrayal toward them if they do not perform, intuitively this does not seem like a central case of trust. It is *mere reliance*, a simple prediction that a person will behave a certain way, not trust.

In its refined form, then, trust seems to have a moral dimension. Philosophers have differed in how they describe this moral component, however. Some philosophers have characterized it as a requirement that there should be mutually shared values between the trusting person and the trusted person (McLeod 2002). Others have suggested a weaker condition than that of shared moral values, since it seems that I can trust somebody even when I know them to have rather different values and standards than myself. It has been variously suggested that in trusting I commit myself to blaming a trusted person if he or she does not perform (Holton 1994), holding them responsible for performing (Walker 2006), or taking a moral expectation toward them (Nickel 2007). None of these suggestions require that the trusted person share the values of the trusting person, only that they be sufficiently responsive to moral requirements and situational expectations that they are suited to be the objects of moral attitudes.

If we think of trust as a morally loaded attitude, what does this mean for the relationship between trust and risk? On this view, trust presupposes an awareness of moral expectations and responsibilities in the trusted person, making that person more trustworthy and reducing the

human risks in interpersonal reliance. If Q, from our earlier example, has as part of her trust a moral expectation of Ron, the content of which is that he keeps information about the group confidential, this presupposes a view of Ron on which he is capable of responding to the morally salient features of the situation, including the vulnerable situation of Q and her group. Thus, if Q is right in her moral expectations, this minimizes the purely interpersonal, strategic risk in relying on Ron, allowing them jointly to focus their attention on their other goals within their cooperative activity, including the avoidance of other risks.

On such a view, one would often expect to see extensive risk evaluation of a person during the process in which trust is established, but once it has been established this activity would largely cease, allowing one to turn one's attention elsewhere. (However, in some relationships, e.g., parent-child relationships, trust will simply be presupposed rather than established.) Continuing surveillance is therefore a sign that trust has not yet been achieved or that it has failed (Baier 1986). This partially captures the first intuitive point discussed earlier in section ➤ [Two Relationships Between Trust and Risk](#), according to which extensive scrutiny of risks is incompatible with trust. However, it also leaves room for relationships in which full trust has yet to be established or is never reached, so that risk evaluation is still ongoing or there are persistent feelings of vulnerability and associated anxiety.

Our second intuition above was that when the risks are greater, trust is more difficult to justify. On moral conceptions of trust, the attitude of trust is partly justified by one's reasons for thinking that the trusted person has certain moral qualities, or is capable of responding to moral expectations. We could therefore look to this feature to explain why trust is more difficult to justify under conditions of risk. Social psychology has shown that people's tendency to do the virtuous or morally required thing *toward a person unknown to them* is highly susceptible to background conditions such as time pressure (Darley and Batson 1973). This suggests that under adverse conditions, moral responsibility is not usually sufficient to generate altruistic or cooperative behavior. Highly risky situations may also create circumstances in which people's moral characteristics are less reliable toward unfamiliar persons. Thus trust in strangers would be difficult to justify on the basis of attributions of moral responsibility. On the other hand, trust toward intimates or people within one's community might be much better justified under such conditions. The moral dispositions of close associates might be expected to "kick in" especially under conditions of risk and adversity, making them even more trustworthy under those conditions. Thus, here too, the moral view of trust only partly supports our earlier intuitive association between risk and trust. (It is important to note here that there are striking instances in which trust in strangers under highly adverse conditions also turns out to be justified, e.g., those discussed in Monroe 1996.)

## Scientific Approaches to Trust and Risk

---

The three conceptual approaches to trust discerned above (trust as interpersonal staking, trust as nonrational, and trust as moral attitude) are also useful for characterizing and subdividing the scientific discussion concerning the development and evolution of trust. In what follows, we reconsider these three approaches, looking at the scientific research carried out within each paradigm. The consistency and explanatory interest of this scientific research may help to settle which approach is correct, or it may imply that we should take a pluralistic view of the nature of trust for empirical purposes.

## The Science of Trust as Rational Risk Taking

Trust as risk taking under uncertainty is at the heart of many game-theoretic approaches to trust: one tries to determine the factors inhibiting and enabling the evolution of trust, given *calculative self-interested agents*. For instance, many scholars have used prisoner's dilemmas as a way of studying trust. In the classic prisoner's dilemma, two players can choose between two moves, either "cooperate" or "defect". Each player gains when both cooperate. But if only one of them cooperates, the other one, the defector, will gain *more*. If both defect, both lose, but not as much as when an agent's cooperative act is defected by the other. Since there is a risk of greater expected loss in case of cooperation (namely, when the other player defects), the best strategy for securing one's own safety is to defect – but this is a suboptimal strategy in terms of collective benefit. In this game, it is in the interest of the individual to forego cooperation, and to defect. The aim, then, is to identify the conditions under which the payoffs of cooperation start to outweigh the payoffs of individual action. If cooperation does emerge, the implication from cooperation to trust seems straightforward: one can expect an agent to cooperate only when she trusts the other to reciprocate her cooperative act.

Morton Deutsch was one of the first to study these ideas empirically and systematically (Deutsch 1960). He looked at how individual trust orientations (cooperative, individualistic, or competitive) affected the likelihood of individuals cooperating, and found that individuals with a cooperative general trust orientation were also more likely to make cooperative choices in prisoners dilemma situations. Deutsch used standard two-player nonzero-sum games, in which players were requested to imagine they were oriented either cooperatively (before playing, players were asked to consider themselves to be partners), individualistically (players were asked to play just to make as much money as possible for themselves), or competitively (players were asked to play just to make as much money as possible for themselves and also to do better than the other). Deutsch found that cooperation was most likely to develop when both players' trust orientations were cooperative.

The main problem with Deutsch's approach (and that of his followers) is that it does not say much about the conditions under which trust develops, unless one takes the cooperative behavior of a person as an indication of that person's disposition to trust, as Deutsch did. Under that condition, however, the exercise becomes circular, since the measure of trust (cooperative behavior) is exactly the feature that trust was said to cause (Cook and Cooper 2003). In other words, when an agent makes a choice in a prisoner's dilemma, she does two things: she decides whether or not to trust the other, and simultaneously, whether to honor the possible trust placed in her. Consequently, her decision to cooperate does not need to reflect her trust, for it might just as well indicate her willingness to reciprocate the other's possible act of trust.

For these reasons, scholars started to look for new game-theoretic protocols in which trust and reciprocity could be dissociated. One such protocol, currently widely used in economics and psychology, is the so-called trust-honor game (see, e.g., Snijders and Keren 1999). Basically it is a sequential prisoner's dilemma, with players no longer making decisions simultaneously, but consecutively. First, player 1 has to decide whether or not to trust player 2, without knowing what the latter will decide; second, player 2, with knowledge of the first player's decision, decides whether to honor the trust by reciprocating.

To get an idea of how trust-honor games work, consider the so-called investment game, developed by Berg et al. (1995). Two players get \$10 each. The game consists of two stages: the trust *placing* and the trust *honoring* stage. In the first stage, player 1 needs to decide how much

money to pass to an anonymous player 2. All the money which is passed is increased by a factor greater than 1, say, 3. So if player 1 passes \$10, player 2 will receive \$30, resulting in the allocation (\$0, \$40). In the second stage, player 2 needs to decide whether or not to honor the trust placed in her: she gets the opportunity either to keep all the money or to send an amount  $x$  back to player 1. For all  $x$ 's smaller than 30 but greater than 10, both players will be better off than in the original allocation (\$10, \$10). For an increase in payoff for either player to occur at all, however, the first stage of the game is crucial: player 1 needs to be willing to take a risk on player 2.

In the study of Berg et al. (1995), players were not only anonymous, their interaction was one shot (taking away the possible effects of reputation building), and there was no opportunity for them to punish dishonored trust. Notwithstanding these conditions, the researchers found that their subjects did both place trust (by sending money) and honor trust (by sending money back). Trust is calculated with the likelihood of reciprocity in mind, and since reciprocity is assumed by the trustor “as a basic element of human behavior” this results in his extending trust even to an anonymous counterpart (p. 122). The reciprocity of an anonymous partner is interesting because one cannot use similarity or other identifying characteristics to gauge her likely future performance. The trustor's reason for positing reciprocation depends on the mere idea of the game situation or of an agent in that situation. Follow-up studies changed the incentive structure of the investment game, added mechanisms of reputation building and punishment, made the encounters between players non-anonymous, looked at cultural background effects, and so on, but one premise remained constant: trust is measured by looking at the *risk* a player is willing to take on the other. In the setup above, player 1's behavior would indicate a high level of trust in case she passed \$10 to player 2, a low level in case she did not pass any money. Thus, Cook and Cooper (2003) write, “First, I decide whether or not to ‘trust’ you (or take a risk on you) and cooperate on the first play of the game, then you must decide whether or not to ‘honor’ that trust by cooperating in turn. Clearly, this behavior can be viewed as risk taking by the first player” (p. 217).

So in these studies, an explanation of trust simply *is* an explanation of the conditions under which subjects decide in favor of social risk taking. Social risk taking, in turn, is calculative, as argued in section [Trust as Interpersonal Staking](#): one takes a risk (i.e., one trusts) in the hope of benefiting afterward (receiving more than the \$10 one started with). If trust is a strategy of generating higher payoffs for individuals, and if these individuals are calculative utility maximizers, trust games would offer a reasonable explanation indeed for why in such essentially self-interested agents as humans, an (apparent) prosocial trait such as trust (or cooperation, more generally) could have evolved.

It is worth briefly discussing the selective pressure commonly invoked here – something which will prove useful for understanding section [The Science of Trust as a Moral Attitude](#). The prehistorical background against which the evolution of cooperation is usually set is a world that was becoming increasingly harsh. In particular, early humans started to cooperate to cope with (1) increased climatic and seasonal variability (Potts 1998; Ash and Gallup 2007); (2) the colonization of new environments, requiring new, cooperative modes of foraging (Kaplan et al. 2000); and/or (3) increased intraspecies competition, given the expansion of hominin populations (Humphrey 1976; Alexander 1989; Dunbar 1998). In these challenging environments, an individual's self-interest was best served by the individual's contributing toward the interest of the group. From a selective point of view, cooperation succeeded where individual efforts began to fail.

The prime cooperative interaction to be explained in these evolutionary scenarios is between genetically *unrelated* agents, namely, how unrelated agents began to form bonds to settle down and defend new environments, or how some 1.5 million years ago hunter *Q* started to trust hunter *R* not to comprise *Q*'s life by dropping out of the hunting coalition early, and to do his fair share of stone throwing and clubbing. In other words, it is assumed that genetically *related* individuals have a natural incentive to cooperate, and that therefore, non-kin cooperation has explanatory prevalence over cooperation among kin. Although this methodological choice has yielded neat models and explanations of why trust and other prosocial behaviors *might* have developed in humans, it has obscured some of the non-calculative aspects of trust, which are discussed in the next two subsections.

In summary, in models that assume utility maximizing agents, it is common to treat trust simply as the willingness of an agent to take a risk on another agent. The aim, then, is to establish the conditions under which these calculative agents may develop such calculative form of trust. The prime explanandum of these models is economic cooperation between unrelated agents, while neglecting cooperation among kin.

## The Science of Trust as Non-Calculative

There are two strands of empirical research which question the strong link between trust and risk taking as pictured in staking accounts. The first relates to a remark made in our conceptual discussion of trust as interpersonal staking (section [Trust as Interpersonal Staking](#)), namely that, contrary to what staking accounts would predict, extensive risk reduction (e.g., through extensive risk assessment) should decrease rather than increase levels of trust. Empirical evidence, now, indeed suggests that if one takes away the risk of defection, one takes away the substrate for trust.

In empirical research, risk reduction is usually introduced by letting agents make formal binding agreements about the transaction in which they are about to engage (see, e.g., Molm et al. [2000, 2009](#)). These are often called *negotiated* exchanges: one negotiates the conditions of the transaction, and secures them through some binding agreement (as is the case in most contemporary market exchanges). Both partners know in advance what each will give and get; the agreement ensures that both partners live up to their promise. These exchanges contrast with so-called *reciprocal* exchanges. The latter are riskier in that partners, at the moment of performing an altruistic act, do not know whether the act will be reciprocated in the future. The altruistic act, one hopes, will elicit altruism in the other, although there is no guarantee to that effect.

Molm and colleagues did not use a behavioral measure of trust (e.g., cooperative behavior), but rather an attitudinal one. That is, they asked subjects about their general commitment and trust toward their exchange partner. They found that trust is more likely to develop in reciprocal than negotiated exchange. In negotiated exchange, one might be *confident* that the other will do as promised, yet remain skeptical about the other's trustworthiness. In reciprocal exchange, by contrast, trustworthiness is tested through real interaction; one gets the opportunity to infer the other's trustworthiness from her behavior rather than from the promises she makes. While interacting, partners typically come to trust each other more, evaluate them more positively, and feel more committed to the relationship. To be sure, both negotiated and reciprocal exchange may yield a fair reallocation of goods. Nonetheless, they have different extra payoffs: increased levels of certainty versus increased levels of trust.

Often the contrast is framed somewhat differently, namely, by drawing on the distinction between cognition-based and affect-based trust (see McAllister 1995). On this account, both negotiated and reciprocal exchanges may yield trust, but of different kinds. Negotiated exchange involves cognition-based trust, which is based on beliefs and estimations of a partner's reliability, and which is abandoned in case the partner behaves differently from how she was expected to behave. Conversely, reciprocal exchange results in affect-based trust. Here beliefs about the other's reliability are complemented with affective states of care and personal regard, and often in addition with a sense of forgiveness for occasional instances in which the partner dishonors the trust placed in her.

The second line of empirical research that undermines the idea of trust being merely calculative suggests that risk taking and trust placing rely on two different psychological mechanisms. Note what we should expect if they did not: risk-aversive people should exhibit a lower propensity to engage in relations of trust, and conversely, risk-minded people should be more trust prone (Ben-Ner and Putterman 2001). Several studies have tried to establish this link, but mostly without success (see, e.g., Eckel and Wilson 2004; Ashraf et al. 2007; Fehr 2009; Ben-Ner and Halldorsson 2010). For instance, in the most comprehensive study, Ben-Ner and Halldorsson (2010) looked for correlations between several measures of trust and several measures of risk orientation. Measures of trust included both behavioral measures (i.e., how subjects behaved in an investment game) and so-called survey measures (determined by means of a questionnaire), indicating a subject's general sense of trust in other people. Also with respect to risk attitudes, both behavioral and survey measures were used. Ben-Ner and Halldorsson found that no measure of risk attitude correlated with any of the measures of trust used. (In contrast, gender and personality – extroversion, openness, agreeableness, conscientiousness, and neuroticism – did predict trusting behavior.)

That risk taking and trust building depend on different psychological mechanisms is also suggested by recent studies on the psychological and behavioral effects of certain hormones. Testosterone, e.g., has been shown to increase risk taking (Apicella et al. 2008), but to decrease trust (Bos et al. 2010). Oxytocin has precisely the opposite effect: it decreases risk taking, but increases trust (Kosfeld et al. 2005).

In sum, both strands of empirical research point to a shortcoming in attempts to characterize trust in terms of risk taking: staking accounts largely ignore the profoundly interpersonal and social aspects of trust. There might be good methodological reasons to do so (e.g., to keep models tractable), but it is important to bear in mind that one might miss out on some salient characteristics of trust. Yet, the research described here shares with staking accounts the concern of explaining trust primarily among strangers (or non-kin). That is what sets them apart from the approach considered next.

## The Science of Trust as a Moral Attitude

As we remarked at the end of subsection [The Science of Trust as Rational Risk Taking](#), it is common to argue that human hypersociality evolved to cope with environments that became ever more challenging. On this account, trust is basically a strategy to safeguard one's own fitness indirectly. By enabling cooperation among non-kin, trust increases the fitness of the collective.

Recently, this dominant view has come under attack. Sarah Hrdy (2009), e.g., argues that human cooperative behavior did not evolve in a context of competition (with other groups of

humans), but rather in a context of cooperative childcare among kin and as-if kin. The idea is that the delayed maturity characteristic of humans requires childcare that cannot be provided by mothers alone. A large fraction of the 13 million calories that are needed to rear a modern human from birth to maturity (approx. age 15–16), for instance, is provided not by the mother, but by other caregivers. In contrast to the great apes, human fathers and older siblings invest fairly extensively in their offspring and younger siblings. Even more crucial is the role of grandmothers: they take over when the mother is too busy, when the mother dies early, or when the mother is simply giving birth to the next infant. According to Hawkes, O’Connell, Blurton Jones, and colleagues, this explains the unusual trait of humans to live actively many years after the birth of their final child (Hawkes et al. 1999). Menopause is an adaptation enabling longer childhoods, and in virtue thereof, increased braininess.

Cooperative childcare (or cooperative breeding, as it is often called) is dependent on three factors. First, the child must be able to elicit attention. It must display its vulnerability, and try to solicit the care not only of its mother, but also of more remote caregivers. Second, the mother must be able to delegate some of her work. She must be willing to put her baby’s fate in the hands of others. In other words, she must trust her helpers, and assume their moral qualities, namely, good intentions and selfless motives. In fact, human mothers are unique in this respect. Unlike the other great apes, human mothers do not hold jealously onto their infants, but willingly allow others to hold them. A human mother has a default sense of trust: she assumes – in the absence of counterevidence – that others will not harm her helpless child. If these strangers did harm the baby nonetheless, there would be a clear sense of wrongdoing, backed up by moral or proto-moral attitudes. Third, both the mother and her helpers (grandparents, fathers, siblings, and other as-if kin) must sense and be responsive to the vulnerability and to the particular needs of the infant. Mothers are particularly well-adapted on this score, as they tend to be attuned to the slightest perturbations in the conditions of their young, and to perceive all neonates as attractive (Hrdy 2009). Arguably, such a positive attitude toward babies applies to the human species in general, since it is also commonly found (in varying degrees) in male adults, teens, adolescents, and so forth.

In the view of authors such as Hrdy, the prosocial behavior exhibited in cooperative breeding forms the cornerstone of our morality. In the context of cooperative breeding, early humans developed their capacity to form relations of trust with kin and as-if kin, a capacity which was extended to non-kin only later. In conclusion, from a scientific perspective it seems to make good sense to view trust as an attitude that originates in our dealings with intimates and community members – as expressed by Carolyn McLeod in Section 4.A. It implies a sphere in which even the most valued things, such as one’s child, can be safely entrusted to others. Trust in cooperative breeding is the prototypical form of trust, from which other more peripheral examples of trust derive.

## Further Research

---

In this final section, we set out a few areas for future philosophical research. As suggested in the previous section, there is still empirical research to be done within the conceptual paradigms of trust. “Empirical philosophy” could contribute to this project. Recent philosophical “experiments” have provided us with additional information about people’s intuitive conceptions of intentionality (e.g., Knobe 2003), weakness of will (Mele 2010), causation (Livengood and

Machery 2007), and consciousness (Knobe and Prinz 2008), so there is no principled reason why similar methods could not yield novel insights regarding (folk) conceptions of trust. Such research may eventually suggest that one of the conceptions of trust is more theoretically interesting or useful than the others, or that there is a way of reconciling them. Such work should take up the relationship of trust and risk as an important challenge.

Second, greater philosophical development of some of the underlying conceptual paradigms would be useful to the understanding of human responses to risk. The non-calculative conception of trust stands in need of greater philosophical conceptualization, a point that Möllering also makes (2006, p. 126). Much of the conceptual work on this theory of trust has been carried out by sociologists. Philosophers may also have something to contribute to such an account of trust. It may also be useful to ask whether the moral account can be reconciled with either the calculative or the non-calculative account of trust, explaining how moral considerations make trust more strategically rational, or how they help people to bypass the need for strategic rationality. Perhaps, too, if one distinguished risk and uncertainty from one another more sharply, or drew distinctions between different kinds of risks and uncertainties, the conceptual relation between risk, uncertainty, and trust would become more evident. Another strategy might be to distinguish more sharply between descriptive or conceptual accounts of trust (i.e., concerning when something *is* trust) and normative accounts (i.e., concerning when one *should* trust).

Finally, an area of particular interest in future philosophical work is understanding how trust is related to technological risk. For several reasons, trust in technology has become a key issue in recent years. First of all, as Meijboom (2008, p. 31) points out, food production and other means of satisfying basic welfare have become increasingly large scale, decentralized, and socially and technologically complex. In this context, perception of technological risk is prevalent as a background fear. Ulrich Beck (1992) famously argues that the imposition of technological risks has reached a critical juncture, at which we must reconsider our practices of justifying and imposing such risks. Horror and science fiction films in which technology gives rise to bad consequences are evidence that this fear resonates with the public. Risk is increasingly impersonal. As Renn puts it, “personal experience of risk has been increasingly replaced by information about risks, and individual control over risk by institutional risk management. As a consequence, people rely more than ever on the credibility and sincerity of those from whom they receive information about risk” (2008, p. 222). However, Onora O’Neill suggests that the solution to fears of technological risks does not lie in making more data available to the public. Paradoxically, she points out, greater openness has not created greater trust in the UK health care system, for example (2002a, pp. 72–73; 2002b, pp. 134–135). After all, the public has a difficult time processing information about risks, and does not have time to assess all the data with respect to every technology they could rely upon. These remarks suggest that trust is crucial to understanding public attitudes toward technological risks. It could be that a philosophical approach emphasizing the centrality of trust to the use of technology would help reconcile technocratic and public attitudes about technological risks.

In this chapter, we have shown that although the relationship of trust and risk is still unsettled, and is grounded in underlying disagreement about the nature of trust, that disagreements about this relationship can be given a clear characterization. This may lead to such disagreements eventually being settled by further methodological reflection and empirical study, or it may just mean that different paradigms of trust can be used for different purposes in thinking about risk. In that case, a sketch of the available paradigms for thinking about this relationship, such as we have given here, will be of some value.

## References

- Alexander R (1989) Evolution of the human psyche. In: Mellars P, Stringer C (eds) *The human revolution: behavioural and biological perspectives on the origins of modern humans*. Princeton University Press, Princeton, pp 455–513
- Apicella C, Dreber A, Campbell B, Gray P, Hoffman M, Little A (2008) Testosterone and financial risk preferences. *Evol Hum Behav* 29:384–390
- Ash G, Gallup G (2007) Paleoclimatic variation and brain expansion during human evolution. *Hum Nat* 18:109–124
- Ashraf N, Bohnet I, Piankov N (2007) Decomposing trust and trustworthiness. *Exp Econ* 9:193–208
- Baier A (1986) Trust and antitrust. *Ethics* 96:231–260
- Beck U (1992) Risk society: towards a new modernity (trans: Ritter M). Sage, Thousand Oaks
- Ben-Ner A, Halldorsson F (2010) Trusting and trustworthiness: what are they, how to measure them, and what affects them. *J Econ Psychol* 31:64–79
- Ben-Ner A, Puterman L (2001) Trusting and trustworthiness. *Boston Univ Law Rev* 81:522–551
- Berg J, Dickhaut J, McCabe K (1995) Trust, reciprocity and social history. *Game Econ Behav* 10:122–142
- Boon SD, Holmes JG (1991) The dynamics of interpersonal trust: resolving uncertainty in the face of risk. In: Hinde RA, Groebel J (eds) *Cooperation and prosocial behavior*. Cambridge University Press, Cambridge, pp 190–211
- Bos P, Terburg D, van Honk J (2010) Testosterone decreases trust in socially naïve humans. *Proc Natl Acad Sci* 107:9991–9995
- Buckwalter W (2010) Knowledge isn't closed on saturday: a study in ordinary language. *Rev Philos Psychol* 1:395–406
- Coleman JS (1990) *Foundations of social theory*. Harvard University Press, Cambridge, MA
- Cook K, Cooper R (2003) Experimental studies of cooperation, trust and social exchange. In: Ostrom E, Walker J (eds) *Trust and reciprocity: interdisciplinary lessons from experimental research*. Russell Sage, New York, pp 209–244
- Darley JM, Batson CD (1973) 'From Jerusalem to Jericho': a study of situational and dispositional variables in helping behavior. *J Personal Social Psychol* 27:100–108
- Das TK, Teng B (2004) The risk-based view of trust: a conceptual framework. *J Bus Psychol* 19:85–116
- DeRose K (1992) Contextualism and knowledge attributions. *Philos Phenomenol Res* 52:913–929
- Deutsch M (1960) The effect of motivational orientation upon trust and suspicion. *Hum Relat* 13:123–139
- Deutsch M (1977) *The resolution of conflict: constructive and destructive processes*. Yale, New Haven
- Dunbar R (1998) The social brain hypothesis. *Evol Anthropol* 6:178–190
- Dunn J (1984) The concept of trust in the politics of John Locke. In: Rorty R, Schneewind JB, Skinner Q (eds) *Philosophy in history: essays on the historiography of philosophy*. Cambridge University Press, Cambridge, pp 279–302
- Earle TC, Siegrist M, Gutscher H (2007) Trust, risk perception and the TCC model of cooperation. In: Siegrist M, Earle TC, Gutscher H (eds) *Trust in cooperative risk management: uncertainty and scepticism in the public mind*. Earthscan, London, pp 1–49
- Eckel CC, Wilson RK (2004) Is trust a risky decision? *J Econ Behav Organiz* 55:447–465
- Erikson EH (1950) *Childhood and society*. W.W. Norton, New York
- Faulkner P (2007) On telling and trusting. *Mind* 116:875–902
- Fehr E (2009) The economics of trust. *J Eur Econ Assoc* 7:235–266
- Gagarin M, Woodruff P (eds) (1995) *Early Greek political thought from Homer to the Sophists*. Cambridge University Press, Cambridge
- Gambetta D (1988) Can we trust trust? In: Gambetta D (ed) *Trust: making and breaking cooperative relations*. Basil Blackwell, London, pp 213–237
- Hansson SO (2004) Philosophical perspectives on risk. *Techné* 8:10–35
- Hardin R (1998) Trust in government. In: Braithwaite VA, Levi M (eds) *Trust and governance*. Russell Sage, New York, pp 9–27
- Hardin R (2006) *Trust*. Polity, New York
- Hawkes K, O'Connell J, Blurton Jones N, Alvarez H, Charnov E (1999) Grandmothering, menopause and the evolution of human life histories. *Proc Natl Acad Sci* 95:1336–1339
- Held V (1968) On the meaning of trust. *Ethics* 78:156–159
- Hieronymi P (2008) The reasons of trust. *Australas J Philos* 86:213–236
- Hobbes T (1968) *Leviathan* [orig. 1651] (Macpherson CB ed). Penguin, New York
- Holton R (1994) Deciding to trust, coming to believe. *Australas J Philos* 74:63–76
- Horsburgh HJN (1961) Trust and social objectives. *Ethics* 72:28–40
- Hrdy S (2009) Mothers and others: the evolutionary origins of mutual understanding. Harvard University Press, Cambridge, MA
- Humphrey N (1976) The social function of intellect. In: Bateson P, Hinde R (eds) *Growing points in ethology*. Cambridge University Press, Cambridge

- Jones K (1996) Trust as an affective attitude. *Ethics* 107:4–25
- Kaplan H, Hill K, Lancaster J, Hurtado AM (2000) A theory of human life history evolution: diet, intelligence, and longevity. *Evol Anthropol* 9:156–185
- Knobe J (2003) Intentional action and side effects in ordinary language. *Analysis* 63:190–194
- Knobe J, Prinz J (2008) Intuitions about consciousness: experimental studies. *Phenom Cogn Sci* 7:67–83
- Kosfeld M, Heinrichs M, Zak P, Fischbacher U, Fehr E (2005) Oxytocin increases trust in humans. *Nature* 435:673–676
- Kramer RM, Carnevale PJ (2001) Trust and intergroup negotiation. In: Brown R, Gaertner S (eds) *Blackwell handbook of social psychology: intergroup processes*. Blackwell, London, pp 431–450
- Laslett P (1988) *John Locke, two treatises of government*. Cambridge University Press, Cambridge
- Lewis JD, Weigert A (1985) Trust as a social reality. *Soc Forces* 63:967–985
- Livingood L, Machery E (2007) The folk probably don't think what you think they think: experiments on causation by absence. *Midwest Stud Philos* 31:107–127
- Luhmann N (1979) *Trust and power: two works by Niklas Luhmann*. Wiley, New York
- McAllister D (1995) Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Acad Manage J* 38:24–59
- McLeod C (2002) Self-trust and reproductive autonomy. MIT Press, Cambridge, MA
- McNair EC (2000) The U.S. Army Corps of Engineers dredging operations and environmental research (DOER) program. In: Herbich JB (ed) *Handbook of dredging engineering*. McGraw-Hill Professional, New York, pp 22.01–22.46
- Meijboom FLB (2008) Problems of trust: a question of trustworthiness. Utrecht University, Dissertation
- Mele A (2010) Weakness of will and akrasia. *Philos Stud* 150:391–404
- Möllerling G (2006) Trust: reason, routine, reflexivity. Elsevier, Amsterdam
- Molm L, Takahashi N, Peterson G (2000) Risk and trust in social exchange: an experimental test of a classical proposition. *Am J Sociol* 105:1396–1427
- Molm L, Schaefer D, Collett J (2009) Fragile and resilient trust: risk and uncertainty in negotiated and reciprocal exchange. *Sociol Theory* 27:1–32
- Monroe KR (1996) *The heart of altruism: perceptions of a common humanity*. Princeton University Press, Princeton
- Nagel J (2008) Knowledge ascriptions and the psychological consequences of changing stakes. *Australas J Philos* 86:279–294
- Nickel PJ (2007) Trust and obligation-ascription. *Ethical Theory Moral Pract* 10:309–319
- Nickel PJ (2009) Trust, staking and expectations. *J Theory Soc Behav* 39:345–362
- Nooteboom B (2002) *Trust: forms, foundations functions, failures and figures*. Edward Elgar, Cheltenham
- O'Neill O (2002a) *A question of trust: the BBC Reith lectures 2002*. Cambridge University Press, Cambridge
- O'Neill O (2002b) Autonomy and trust in bioethics. Cambridge University Press, Cambridge
- Pettit P (1995) The cunning of trust. *Philos Public Aff* 24:202–225
- Potts R (1998) Variability selection in hominid evolution. *Evol Anthropol* 7:81–96
- Renn O (2008) *Risk governance: coping with uncertainty in a complex world*. Earthscan, London
- Seligman AB (1997) *The problem of trust*. Princeton University Press, Princeton
- Snijders C, Keren G (1999) Determinants of trust. In: Budescu D, Erev I (eds) *Games and human behavior: essays in honor of Amnon Rapaport*. Lawrence Erlbaum, Mahwah
- Sztompka P (1999) *Trust: a sociological theory*. Cambridge University Press, Cambridge
- Thomas DO (1978) The duty to trust. *Proc Aristotelian Soc* 79:89–101
- Uslaner E (1992) *The moral foundations of trust*. Cambridge University Press, New York
- Verbeek B (2002) *Instrumental rationality and moral philosophy: an essay on the virtues of cooperation*. Kluwer, Dordrecht
- Walker MU (2006) *Moral repair: reconstructing moral relations after wrongdoing*. Cambridge University Press, Cambridge

# 35 Risk and Responsibility

Ibo van de Poel<sup>1</sup> · Jessica Nilén Fahlquist<sup>1,2</sup>

<sup>1</sup>Delft University of Technology, Delft, The Netherlands

<sup>2</sup>Royal Institute of Technology, Stockholm, Sweden

<i>Introduction</i> .....	878
<i>Conceptions of Risk and Responsibility</i> .....	879
Conceptions of Risk .....	879
Conceptions of Responsibility .....	882
Conceptual Relations Between Risk and Responsibility .....	885
<i>Responsibility for Risks</i> .....	886
The Responsibility of Engineers .....	887
Risk Assessment Versus Risk Management .....	888
Individual Versus Collective Responsibility for Risks .....	890
Risk Communication .....	895
<i>Further Research: Organizing Responsibility for Risks</i> .....	897
The Problem of Many Hands (PMH) .....	897
Climate Change as an Example .....	900
Responsibility as a Virtue .....	901
The Procedure of Responsibility Distribution .....	902
Institutional Design .....	904
<i>Conclusion</i> .....	905

**Abstract:** When a risk materializes, it is common to ask the question: who is responsible for the risk being taken? Despite this intimate connection between risk and responsibility, remarkably little has been written on the exact relation between the notions of risk and responsibility. This contribution sets out to explore the relation between risk and responsibility on basis of the somewhat dispersed literature on the topic and it sketches directions for future research. It deals with three more specific topics. First we explore the conceptual connections between risk and responsibility by discussing different conceptions of risk and responsibility and their relationships. Second, we discuss responsibility for risk, paying attention to four more specific activities with respect to risks: risk reduction, risk assessment, risk management, and risk communication. Finally, we explore the problem of many hands (PMH), that is, the problem of attributing responsibility when large numbers of people are involved in an activity. We argue that the PMH has especially become prominent today due to the increased collective nature of actions and due to the fact that our actions often do not involve direct harm but rather risks, that is, the possibility of harm. We illustrate the PMH for climate change and discuss three possible ways of dealing with it: (1) responsibility-as-virtue, (2) a procedure for distributing responsibility, and (3) institutional design.

## Introduction

---

Risk and responsibility are central notions in today's society. When the Deepwater Horizon oil rig exploded in April 2010 killing 11 people and causing a major oil spill in the Gulf of Mexico, questions were asked whether no unacceptable risks had been taken and who was responsible. The popular image in cases like this appears to be that if such severe consequences occur, someone must have, deliberately or not, taken an unacceptable risk and for that reason that person is also responsible for the outcome. One reason why the materialization of risks immediately raises questions about responsibility is our increased control over the environment. Even in cases of what are called natural risks, that is, risks with primarily natural rather than human causes, questions about responsibility seem often appropriate nowadays. When an earthquake strikes a densely populated area and kills thousands of people, it may be improper to hold someone responsible for the mere fact that the earthquake occurred, but it might well be appropriate to hold certain people responsible for the fact that no proper warning system for earthquakes was in place or for the fact that the buildings were not or insufficiently earthquake resistant. In as far as both factors mentioned contributed to the magnitude of the disaster, it might even be appropriate to hold certain people responsible for the fatalities.

The earthquake example shows that the idea that it is by definition impossible to attribute responsibility for natural risks and that such risks are morally less unacceptable is increasingly hard to maintain, especially due to technological developments. This may be considered a positive development in as far it has enabled mankind to drastically reduce the number of fatalities, and other negative consequences, as a result of natural risks. At the same time, technological development and the increasing complexity of society have introduced new risks; the Deepwater Horizon oil rig is just one example. Especially in the industrialized countries, these new risks now seem to be a greater worry than the traditional so-called natural risks. Although these new risks are clearly man-made, they are in practice not always easy to control. It is also often quite difficult to attribute responsibility for them due to the larger number of people involved; this is sometimes referred to as the "problem of many hands" (PMH), which we will describe and analyze in more detail in section ➤ [Further Research: Organizing Responsibility for Risks](#). Before we do

so, we will first discuss the relation between risk and responsibility on a more abstract, conceptual level by discussing different conceptions of risk and responsibility and their relation, in section [Conceptions of Risk and Responsibility](#). Section [Responsibility for Risks](#) focuses on the responsibility for dealing with risk; it primarily focuses on so-called forward-looking moral responsibility and on technological risks.

While risk and responsibility are central notions in today's society and a lot has been written about both, remarkably few authors have explicitly discussed the relation between the two. Moreover, the available literature is somewhat dispersed over various disciplines, like philosophy, sociology, and psychology. As a consequence, it is impossible to make a neat distinction between the established state of the art and future research in this contribution. Rather the contribution as a whole has a somewhat explorative character. Nevertheless, sections [Conceptions of Risk and Responsibility](#) and [Responsibility for Risks](#) mainly discuss the existing literature, although they make some connections that cannot be found in the current literature. Section [Further Research: Organizing Responsibility for Risks](#) explores the so-called problem of many hands and the need to organize responsibility, which is rather recent and requires future research, although some work has already been done and possible directions for future research can be indicated.

## Conceptions of Risk and Responsibility

---

Both risk and responsibility are complex concepts that are used in a multiplicity of meanings or conceptions as we will call them. Moreover, as we will see below, while some of these conceptions are merely descriptive, others are clearly normative. Before we delve deeper into the relation between risk and responsibility, it is therefore useful to be more precise about both concepts. We will do so by first discussing different conceptions of risk (section [Conceptions of Risk](#)) and of responsibility (section [Conceptions of Responsibility](#)). We use the term “conception” here to refer to the specific way a certain concept like risk or responsibility is understood. The idea is that while different authors, approaches, or theories may roughly refer to the same concept, the way they understand the concept and the conceptual relations they construe with other concepts is different. After discussing some of the conceptions of risk and responsibility, section [Conceptual Relations Between Risk and Responsibility](#) discusses conceptual relations between risk and responsibility.

### Conceptions of Risk

---

The concept of risk is used in different ways (see [Chap. 3, The Concepts of Risk and Safety](#), in this handbook). Hansson (2009, pp. 1069–1071), for example, mentions the following conceptions:

1. Risk = an *unwanted event* that may or may not occur
2. Risk = the *cause* of an unwanted event that may or may not occur
3. Risk = the *probability* of an unwanted event that may or may not occur
4. Risk = the statistical expectation value of an unwanted event that may or may not occur
5. Risk = the fact that a decision is made under conditions of *known probabilities* (“decision under risk”)

The fourth conception has by now become the most common technical conception of risk and this conception is used usually in engineering and in risk assessment. The fifth conception is common in decision theory. In this field, it is common to distinguish decisions under risk from decisions under certainty and decisions under uncertainty. Certainty refers to the situation in which the outcomes (or consequences) of possible actions are certain. Risk refers to the situation in which possible outcomes are known and the probabilities (between 0 and 1) of occurrence of these outcomes are known. Uncertainty refers to situations in which possible outcomes are known but no probabilities can be attached to these outcomes. A situation in which even possible outcomes are unknown may be referred to as ignorance.

The fifth, decision-theoretical conception of risk is congruent with the fourth conception in the sense that both require knowledge of possible outcomes and of the probability of such outcomes to speak meaningfully about a risk. One difference is that whereas the fifth conception does not distinguish between wanted and unwanted outcomes, the fourth explicitly refers to unwanted outcomes. Both the fourth and the fifth conception are different from the way the term “risk” is often used in daily language. In daily language, we commonly refer to an undesirable event as a risk, even if the probability is unknown or the exact consequences are unknown. One way to deal with this ambiguity is to distinguish between hazards (or dangers) and risks. Hazard refers to the mere possibility of an unwanted event (conception 1 above), without necessarily knowing either the consequences or the probability of such an unwanted event. Risk may then be seen as a specification of the notion of hazard. The most common definition of risk in engineering and risk assessment, and more generally in techno-scientific contexts, is that of statistical expectation value, or the product of the consequences of an unwanted event and the probability of the unwanted event occurring (meaning 4 above). But even in techno-scientific contexts other definitions of risk can be found. The International Program on Chemical Safety, for example, in an attempt to harmonize the different meanings of terms used in risk assessment defines risk as: “The probability of an adverse effect in an organism, system, or (sub)population caused under specified circumstances by exposure to an agent” (International Program on Chemical Safety 2004, p. 13). This is closer to the third than the fourth conception mentioned by Hansson. Nevertheless, the International Program on Chemical Safety appears to see risk as a further specification of hazard, which they define as: “Inherent property of an agent or situation having the potential to cause adverse effects when an organism, system, or (sub)population is exposed to that agent” (International Program on Chemical Safety 2004, p. 12).

Conceptions of risk cannot only be found in techno-scientific contexts and in decision theory, but also in social science, in literature on risk perception (psychology), and more recently in moral theory (for a discussion of different conceptions of risk in different academic fields, see Bradbury 1989; Thompson and Dean 1996; Renn 1992; Shrader-Frechette 1991a). We will below discuss some of the main conceptions of risk found in these bodies of literature. The technical conception of risk assumes, at least implicitly, that the only relevant aspects of risk are the magnitude of certain unwanted consequences and the probability of these consequences occurring. The conception nevertheless contains a normative element because it refers to *unwanted* consequences (or events). However, apart from this normative element, the conception is meant to be descriptive rather than normative. Moreover, it is intended to be context free, in the sense that it assumes that the only relevant information about a risky activity is the probability and magnitude of consequences (Thompson and Dean 1996). Typically, conceptions of risk in psychology, social science, and moral theory are more

contextual. They may refer to such contextual information as by whom the risk is run, whether the risk is imposed or voluntary, whether it is a natural or man-made risk, and so on. What contextual elements are included, and the reason for which contextual elements are included is, however, different for different contextual conceptions of risk.

The psychological literature on risk perception has established that lay people include contextual elements in how they perceive and understand risks (e.g., Slovic 2000). These include, for example, dread, familiarity, exposure, controllability, catastrophic potential, perceived benefits, time delay (future generations), and voluntariness. Sometimes the fact that lay people have a different notion of risk than experts, and therefore estimate the magnitude of risks differently, is seen as a sign of their irrationality. This interpretation assumes that the technical conception of risk is the right one and that lay people should be educated to comply with it. Several authors have, however, pointed out that the contextual elements included by lay people are relevant for the acceptability of risks and for risk management and that in that sense the public's conception of risk is "richer" and in a sense more adequate than that of scientific experts (e.g., Slovic 2000; Roeser 2006, 2007). In the literature on the ethics of risk it is now commonly accepted that the moral acceptability of risks depends on more concerns than just the probability and magnitude of possible negative consequences (see [Chap. 30, Ethics and Risk](#), in this handbook). Moral concerns that are often mentioned include voluntariness, the balance and distribution of benefits and risks (over different groups and over generations), and the availability of alternatives (Asveld and Roeser 2009; Shrader-Frechette 1991b; Hansson 2009; Harris et al. 2008; Van de Poel and Royakkers 2011).

In the social sciences, a rich variety of conceptions of risk have been proposed (Renn 1992) (see [Chap. 40, Sociology of Risk](#), in this handbook). We will not try to discuss or classify all these conceptions, but will briefly outline two influential social theories of risk, that is, cultural theory (Douglas and Wildavsky 1982) and risk society (Beck 1992). Cultural theory conceives of risks as collective, cultural constructs (see [Chap. 28, Cultural Cognition as a Conception of the Cultural Theory of Risk](#), in this handbook). Douglas and Wildavsky (1982) distinguish three cultural biases that correspond to and are maintained by three types of social organization: hierarchical, market individualistic, and sectarian. They claim that each bias corresponds to a particular selection of dangers as risks. Danger here refers to what we above called a hazard: the (objective) possibility of something going wrong. According to Douglas and Wildavsky, dangers cannot be known directly. Instead they are culturally constructed as risks. Depending on the cultural bias, certain dangers are preeminently focused on. Hierarchists focus on risks of human violence (war, terrorism, and crime), market individualists on risks of economic collapse, and sectarians on risks of technology (Douglas and Wildavsky 1982, pp. 187–188).

Like Douglas and Wildavsky, Ulrich Beck in his theory of risk society sees risk as a social construct. But whereas Douglas and Wildavsky focus on the cultural construction of risks and believe that various constructions may exist side by side, Beck places the social construction of risk in historical perspective. Beck defines risk as "*a systematic way of dealing with hazards and insecurities induced and introduced by modernization itself*" (Beck 1992, p. 21, emphasis in the original). Speaking in terms of risks, Beck claims, is historically a recent phenomenon and it is closely tied to the idea that risks depend on decisions (Beck 1992, p. 183). Typically for what Beck calls the "risk society" is that it has become impossible to attribute hazards to external causes. Rather, all hazards are seen as depending on human choice and, hence, are, according to Beck's definition of these notions, conceived as risks. Consequently, in risk society the

central issue is the allocation of risk rather than the allocation of wealth as it was in industrial society.

Some authors have explicitly proposed to extend the technical conception of risk to include some of the mentioned contextual elements. We will briefly outline two examples. Rayner (1992) has proposed the following adaption to the conventional conception of risk:

$$R = (P \times M) + (T \times L \times C)$$

with

R = Risk

P = Probability of occurrence of the adverse event

M = Magnitude of the adverse consequences

T = Trustworthiness of the institutions regulating the technology

L = Acceptability of the principle used to apportion liabilities for undesired consequences

C = Acceptability of the procedure by which collective consent is obtained to those who must run the consequences

Although this conception has a number of technical difficulties, it brings to the fore some of the additional dimensions that are important not just for the perception or cultural construction of risks but also for their regulation and moral acceptability.

More recently, Wolff (2006) has proposed to add cause as a primary variable in addition to probability and magnitude to the conception of risk. The rationale for this proposal is that cause is also relevant for the acceptability of risks. Not only may there be a difference between natural and man-made risks, but also different man-made risks may be different in acceptability depending on whether the human cause is based on culpable or non-culpable behavior and the type of culpable behavior (e.g., malice, recklessness, negligence, or incompetence). We might have good moral reasons to consider risks based on malice (e.g., a terrorist attack) less acceptable than risks based on incompetence even when they are roughly the same in terms of probability and consequences. In addition to cause, Wolff proposes to add such factors as fear (dread), blame, and shame as secondary variables that might affect each of the primary variables. Like in the case of Rayner's conception, the technicalities of the new conception are somewhat unclear, but it is definitively an attempt to broaden the conception of risk to include contextual elements that are important for the (moral) acceptability of risks.

Rayner's and Wolff's proposals raise the question whether all factors which are relevant for decisions about acceptable risk or risk management should be included in a conceptualization of risk. Even if it is reasonable to include moral concerns in our decisions about risks, it may be doubted whether the best way to deal with such additional concerns is to build them into a (formal) conception of risk.

## Conceptions of Responsibility

Like the notion of risk, the concept of responsibility can be conceptualized in different ways. One of the first authors to distinguish different conceptions of responsibility was Hart (1968, pp. 210–237) who mentions four main conceptions of responsibility: role responsibility, causal responsibility, liability responsibility, and capacity responsibility. Later authors have

distinguished additional conceptions, and the following gives a good impression of the various conceptions that might be distinguished (Van de Poel 2011):

1. Responsibility-as-cause. As in: the earthquake caused the death of 100 people.
2. Responsibility-as-role. As in: the train driver is responsible for driving the train.
3. Responsibility-as-authority. As in: he is responsible for the project, meaning he is in charge of the project. This may also be called responsibility-as-office or responsibility-as-jurisdiction. It refers to a realm in which one has the authority to make decisions or is in charge and for which one can be held accountable.
4. Responsibility-as-capacity. As in: the ability to act in a responsible way. This includes, for example, the ability to reflect on the consequences of one's actions, to form intentions, to deliberately choose an action and act upon it.
5. Responsibility-as-virtue, as the disposition (character trait) to act responsibly. As in: he is a responsible person. (The difference between responsibility-as-capacity and responsibility-as-virtue is that whereas the former only refers to ability, the second refers to a disposition that is also surfacing in actions. So someone who has the capacity for responsibility may be an irresponsible person in the virtue sense).
6. Responsibility-as-obligation to see to it that something is the case. As in: he is responsible for the safety of the passengers, meaning he is responsible to see to it that the passengers are transported safely.
7. Responsibility-as-accountability. As in: the (moral obligation) to account for what you did or what happened (and your role in it happening).
8. Responsibility-as-blameworthiness. As in: he is responsible for the car accident, meaning he can be properly blamed for the car accident happening.
9. Responsibility-as-liability. As in: he is liable to pay damages.

The first four conceptions are more or less descriptive: responsibility-as-cause, role, authority, or capacity describes something that is the case or not. The other five are more normative. The first two normative conceptions – responsibility-as-virtue and responsibility-as-obligation – are primarily forward-looking (prospective) in nature. Responsibility-as-accountability, blameworthiness, and liability are backward-looking (retrospective) in the sense that they usually apply to something that has occurred. Both the forward-looking and the backward-looking normative conception of responsibility are relevant in relation to risks. Backward-looking responsibility is mainly at stake when a risk has materialized and then relates to such questions like: Who is accountable for the occurrence of the risk? Who can be properly blamed for the risk? Who is liable to pay the damage resulting from the risk materializing? Forward-looking responsibility is mainly relevant with respect to the prevention and management of risks. It may refer to different tasks that are relevant for preventing and managing risk like risk assessment, risk reduction, risk management, and risk communication. We will discuss the responsibility for these tasks in section [● Responsibility for Risks](#).

Cross-cutting the distinction between the different conceptions of responsibility is a distinction between what might be called different “sorts” of responsibility like organizational responsibility, legal responsibility, and moral responsibility. The main distinction between these sorts is the grounds on which it is determined whether someone is responsible (in one of the senses distinguished above). Organizational responsibility is mainly determined by the rules and roles that exist in an organization, legal responsibility by the law (including

jurisprudence), and moral responsibility is based on moral considerations. The two types of distinctions are, however, not completely independent of each other. Organizational responsibility, for example, often refers to responsibility-as-task or responsibility-as-authority and seems unrelated to responsibility-as-cause and responsibility-as-capacity. It might also refer to responsibility-as-accountability, just like legal and moral responsibility. We might thus distinguish between organizational, legal, and moral accountability, where the first is dependent on an organization's rules and roles, the second on the law, and the third on moral considerations.

In this contribution we mainly focus on moral responsibility. Most of the general philosophical literature on responsibility has focused on backward-looking moral responsibility, in particular on blameworthiness. In this literature also, a number of general conditions have been articulated which should be met in order for someone to be held properly or fairly responsible (e.g., Wallace 1994; Fischer and Ravizza 1998). Some of these conditions, especially the freedom and knowledge condition, go back to Aristotle (*The Nicomachean Ethics*, book III, Chaps. 1–5). These conditions include:

1. Moral agency. The agent A is a moral agent, that is, has the capacity to act responsibly (responsibility-as-capacity).
2. Causality. The agent A is somehow causally involved in the action or outcome for which A is held responsible (responsibility-as-cause).
3. Wrongdoing. The agent A did something wrong.
4. Freedom. The agent A was not compelled to act in a certain way or to bring about a certain outcome.
5. Knowledge. The agent A knew, or at least could reasonably have known that a certain action would occur or a certain outcome would result and that this was undesirable.

Although these general conditions can be found in many accounts, there is much debate about at least two issues. One is the exact content and formulation of each of the conditions. For example, does the freedom condition imply that the agent could have acted otherwise (e.g., Frankfurt 1969)? The other is whether these conditions are individually necessary and jointly sufficient in order for an agent to be blameworthy. One way to deal with the latter issue is to conceive of the mentioned conditions as arguments or reasons for holding someone responsible (blameworthy) for something rather than as a strict set of conditions (Davis *forthcoming*).

Whereas the general philosophical literature on responsibility has typically focused on backward-looking responsibility, the more specific analyses of moral responsibility in technoscientific contexts, and more specifically as applied to (technological) risks, often focus on forward-looking responsibility. They, for example, discuss the forward-looking responsibility of engineers for preventing or reducing risks (e.g., Davis 1998; Harris et al. 2008; Martin and Schinzingher 2005; Van de Poel and Royakkers 2011). One explanation for this focus may be that in these contexts the main aim is to prevent and manage risks rather than to attribute blame and liability. This is, of course, not to deny that in other contexts, backward-looking responsibility for risks is very relevant. It surfaces, for example, in court cases about who is (legally) liable for certain damage resulting from the materialization of technological risks. It is also very relevant in more general social and political discussions about how the costs of risks should be borne: by the victim, by the one creating the risks, or collectively by society, for example, through social insurance.

## Conceptual Relations Between Risk and Responsibility

The conceptual connections between risk and responsibility depend on which conception of risk and which conception of responsibility one adopts. The technical conception of risk, which understands risks as the product of probability and magnitude of certain undesirable consequences, is largely descriptive, but it contains a normative element because it refers to undesirable outcomes. Typically, responsibility also is often used in reference to undesirable outcomes, especially if responsibility is understood as blameworthiness. Yet if the undesirable consequences, to which the technical conception of risk refers, materialize this does not necessarily imply that someone is blameworthy for these consequences. As we have seen, a number of conditions have to be met in order for someone to fairly be held responsible for such consequences. In cases of risks the knowledge condition will usually be fulfilled because if a risk has been established it is known that certain consequences might occur. It will often be less clear whether the wrongdoing condition is met. Risks normally refer to unintended, but not necessarily unforeseen, consequences of action. Nevertheless, under at least two circumstances, the introduction of a risk amounts to wrongdoing. One is if the actor is reckless, that is, if he knows that a risk is (morally) unacceptable but still exposes others to it. The other is negligence. In the latter case, the actor is unaware of the risk he is taking but should and could have known the risk and exposing others to the risk is unacceptable.

If we focus on forward-looking responsibility rather than backward-looking responsibility, the technical conception of risk might be thought to imply an obligation to avoid risks since most conceptions of risk refer to something undesirable. Again, however, the relation is not straightforward. Some risks, like certain natural risks, may be unavoidable. Other risks may not be unavoidable but worth taking given the advantages of certain risky activities. Nevertheless, there seems to be a forward-looking responsibility to properly deal with risks. In section [❸ Responsibility for Risks](#), we will further break down that responsibility and discuss some of its main components.

In the psychological literature on risk perception, no direct link is made between risk and responsibility. Nevertheless, some of the factors that this literature has shown to influence the perception of risk may be linked to the concept of responsibility. One such factor is controllability (e.g., Slovic 2000). Control is often seen as a precondition for responsibility; it is linked to the conditions of freedom and knowledge we mentioned above. Also voluntariness, another important factor in the perception of risk (e.g., Slovic 2000), is linked to those responsibility conditions. This suggests that risks for which one is not responsible (or cannot take responsibility) but to which one is exposed beyond one's will and/or control are perceived as larger and less acceptable.

In the sociological literature on risk that we discussed in section [❸ Conceptions of Risk](#), a much more direct connection between risk and responsibility is supposed. Mary Douglas (1985) argues that the same institutionally embedded cultural biases that shape the social construction of risks also shape the attribution of responsibility, especially of blameworthiness. Institutions are, according to Douglas, typically characterized by certain recurring patterns of attributing blame, like blaming the victim, or blaming outsiders, or just accepting the materialization of risks as fate or the price to be paid for progress. According to the theory of risk society, both risk and responsibility are connected to control and decisions. This implies a rather direct conceptual connection between risk and responsibility. As Anthony Giddens expresses it: "The relation between risk and responsibility can be easily stated, at least on an

abstract level. Risks only exist when there are decisions to be taken . . . The idea of responsibility also presumes decisions. What brings into play the notion of responsibility is that someone takes a decision having discernable consequences” (Giddens 1999, p. 8). The socio-logical literature seems to refer primarily to organizational responsibility, in the sense that attribution of responsibility primarily depends on social conventions. Nevertheless, as we have seen the idea of control, which is central to risk in the theory of risk society, is also central for moral responsibility.

The redefinitions of risk proposed by Rayner and Wolff, finally, both refer to responsibility as an ingredient in the conception of risk. Rayner includes liability as an aspect of risk. While liability is usually primarily understood as a legal notion, his reference to the *acceptability* of liability procedures also has clear moral connotations. In Wolff’s conception of risk, responsibility affects the variable “cause” that he proposes as additional primary variable for risk. As Wolff points out, it matters for the acceptability of risk whether it is caused by malice, recklessness, or negligence. These distinctions also have a direct bearing on the moral responsibility of the agent causing the undesirable consequences; they represent different degrees of wrongdoing. So, on Wolff’s conceptualization, whether and to what degree anyone is responsible for a risk has a bearing on the acceptability of that risk.

Although the relation between risk and responsibility depends on the exact conceptualization of both terms and one might discuss how to best conceptualize both terms, the above discussion leads to a number of general conclusions. First, if an undesired outcome is the result of someone taking a risk or exposing others to a risk, it appears natural to talk about responsibility in the backward-looking sense (accountability, blameworthiness, liability) for those consequences and for the risk taken. Second, both risk and responsibility are connected to control and decisions. Even if one does not accept the tight conceptual connection between risk and control that the theory of risk society supposes, it seems clear that risks often are related to decisions and control. As pointed out in the introduction, even so-called natural risks increasingly come under human control. This implies that we can not only hold people responsible for risks in a backward-looking way, but that people can also take or assume forward-looking responsibility (responsibility-as-virtue or as obligation) for risks. Third, the acceptability of risks appears to depend, at least partly, on whether someone can fairly be held responsible for the risk occurring or materializing.

## **Responsibility for Risks**

---

In the literature on risk some general frameworks have been developed for thinking about the responsibility for risks and some general tentative answers have been formulated to the question who is responsible for certain risks. In this section, we present a number of these positions and the debates to which they have given rise. We focus on human-induced risks, that is, nonnatural risks, with a prime focus on technological risks. Our focus is also primarily on forward-looking responsibility rather than on backward-looking responsibility (accountability, blameworthiness, and liability) for risks.

Forward-looking responsibility for risks can be subdivided in the following main responsibilities:

1. Responsibility for risk reduction.
2. Responsibility for risk assessment, that is, establishing risks and their magnitude.

3. Responsibility for risk management. Risk management includes decisions about what risks are acceptable and the devising of regulations, procedures, and the like to ensure that risks remain within the limits of what is acceptable.
4. Responsibility for risk communication, that is, the communication of certain risks, in particular to the public.

Section ② [The Responsibility of Engineers](#) will discuss the responsibility for risk reduction. In the case of technological risks, this responsibility is often attributed to engineers. Section ② [Risk Assessment Versus Risk Management](#) will focus on the responsibility for risk assessment versus risk management. The former is often attributed to scientists, while governments and company managers are often held responsible for the latter. It will be examined whether this division of responsibilities is justified. Section ② [Individual Versus Collective Responsibility for Risks](#) will focus on an important issue with respect to risk management: whether decisions concerning acceptable risk are primarily the responsibility of individuals who take and potentially suffer the risk or whether it is a collective responsibility that should be dealt with through regulation by the government. Section ② [Risk Communication](#) will discuss some of the responsibilities of risk communicators and related dilemmas that have been discussed in the literature on risk communication.

## The Responsibility of Engineers

---

Engineers play a key role in the development and design of new technologies. In this role they also influence the creation of technological risks. In the engineering ethics literature, it is commonly argued that engineers have a responsibility for safety (Davis 1998; Harris et al. 2008; Martin and Schinzinger 2005; Van de Poel and Royakkers 2011). In this section, we will consider these arguments and discuss how safety and risk are related and what the engineers' responsibility for safety implies for their responsibility for technological risks.

Most engineering codes of ethics state that engineers have a responsibility for the safety of the public. Thus, the code of the National Society of Professional Engineers in the USA states that: "Engineers, in the fulfillment of their professional duties, shall: . . . Hold paramount the safety, health, and welfare of the public" (NSPE 2007). Safety is not only stressed as the engineer's responsibility in codes of ethics but also in technical codes and standards. Technical codes are legal requirements that are enforced by a governmental body to protect safety, health, and other relevant values (Hunter 1997). Technical standards are usually recommendations rather than legal requirements that are written by engineering experts in standardization committees. Standards are usually more detailed than technical codes and may contain detailed provisions about how to design for safety.

Does the fact that safety is a prime concern in engineering codes of ethics and technical codes and standards entail that engineers have a moral responsibility for safety? One can take different stances here. Some authors have argued that codes of ethics entail an implicit contract either between a profession and society or among professionals themselves. Michael Davis, for example, defines a profession as "a number of individuals in the same occupation voluntarily organized to earn a living by openly serving a certain moral ideal in a morally permissible way beyond what law, market, and morality would otherwise require" (Davis 1998, p. 417). This moral idea is laid down in codes of ethics and thus implies, as we have seen, a responsibility for

safety. According to Davis, codes are binding because they are an implicit contract between professionals, to which engineers subscribe by joining the engineering profession.

One could also argue that codes of ethics or technical codes and standards as such do not entail responsibilities for engineers but that they *express* responsibilities that are grounded otherwise. In that case, the engineers' responsibility for safety may, for example, be grounded in one of the general ethical theories like consequentialism, deontology, or virtue ethics. But if we believe that engineers have a moral responsibility for safety, does this also entail a responsibility for risks? To answer this question, we need to look a bit deeper into the conceptual relation between safety and risk (see ➤ [Chap. 3, The Concepts of Risk and Safety](#), in this handbook). In engineering, safety has been understood in different ways. One understanding is that safety means absolute safety and, hence, implies the absence of risk. In most contexts, this understanding is not very useful (Hansson 2009, p. 1074). Absolute safety is usually impossible and even if it would be possible it would in most cases be undesirable because eliminating risks usually comes at a cost, not only in monetary terms but also in terms of other design criteria like sustainability or ease of use. It is therefore better to understand safety in terms of "acceptable risk." One might then say that a technological device is safe if its associated risks are acceptable. What is acceptable will depend on what is feasible and what is reasonable. The notion of reasonableness refers here to the fact that reducing risks comes at a cost and that hence not all risk reductions are desirable.

So conceived, engineers may be said to be responsible for reducing risks to an acceptable level. What is acceptable, however, requires a normative judgment. This raises the question whether the engineer's responsibility for reducing risks to an acceptable level includes the responsibility to make a normative judgment on which risks are acceptable and which ones are not or that it is limited to meeting an acceptable risk level that is set in another way, for example, by a governmental regulator. The answer to this question may well depend on whether the engineers are designing a well-established technology for which safety standards have been set that are generally and publicly recognized as legitimate or that they are designing a radically new technology, like nanotechnology, for which existing safety standards cannot be applied straightforwardly and of which the hazards and risks are more uncertain anyway (for this distinction, see Van de Poel and Van Gorp 2006). In the former case, engineers can rely on established safety standards. In the latter case, such standards are absent. Therefore in the second case engineers and scientists also have some responsibility for judging what risks are acceptable, although they are certainly not the only party that is or should be involved in such judgments.

## Risk Assessment Versus Risk Management

---

In the previous section we have seen that a distinction needs to be made between responsibility for risk reduction and responsibility for decisions about acceptable risks. Engineers have a responsibility for risk reduction but not necessarily or at least to a lesser degree a responsibility for deciding about acceptable risk. In this section we will discuss a somewhat similar issue in the division of responsibility for risk, namely, the responsibility for establishing the magnitude of risks (risk assessment) and decisions about the acceptability and management of risks (risk management). Traditionally risk assessment is seen as a responsibility of scientists, and risk management as a responsibility of governments and (company) managers (National Research Council 1983) (see ➤ [Chap. 46, Risk Management in Technocracy](#), in this

handbook). In this section, we will discuss whether this division of labor and responsibility is tenable or not. In particular, we will focus on the question whether adequate risk assessment can be completely value free, as is often supposed, or, as has been argued by a number of authors, that it needs to rely on at least some value judgments.

One reason why risk assessment cannot be entirely value free is that in order to do a risk assessment a decision needs to be made on what risks to focus. Since, on the conventional technical conception of risk (see section [Conceptions of Risk](#)), risks are by definition undesirable, classifying something as a risk already involves a value judgment. It might be argued, nevertheless, that decisions about what is undesirable are to be made by risk managers and that risk assessors, as scientists, should then investigate all potential risks. In practice, however, a risk assessment cannot investigate all possible risks; a selection will have to be made and selecting certain risks rather than others implies a value judgment. Again, it can be argued that this judgment is to be made by risk managers. A particular problem here might be that some risks are harder to investigate or establish scientifically than others. Some risks may even be statistically undetectable (Hansson 2009, pp. 1084–1086). From the fact that a risk is hard or even impossible to detect scientifically, of course it does not follow that it is also socially or morally unimportant or irrelevant, as it might have important consequences for society if it manifests itself after all. This already points to a possible tension between selecting risks for investigation from a scientific point of view and from a social or moral point of view.

The science of risk assessment also involves value judgments with respect to a number of methodological decisions that are to be made during risk assessment. Such methodological decisions influence the risk of error. A risk assessment might, due to error, wrongly estimate a certain risk or it might establish a risk where actually none exists. Heather Douglas (2009) argues that scientists in general have a responsibility to consider the consequences of error, just like anybody else. While this may seem common sense, it has important consequences once one takes into account the social ends for which risk assessments are used. Risk assessment is not primarily used to increase the stock of knowledge, but rather as an input for risk management. If a risk assessment wrongly declares something not to be a risk while it actually is a serious risk, or vice versa, this may lead to huge social costs, both in terms of fatalities and economic costs.

Various authors have therefore suggested that, unlike traditional science, risk assessment should primarily avoid what are called type 2 errors rather than type 1 errors (Cranor 1993; Shrader-Frechette 1991b; Hansson 2008; see also Hansson's [Chap. 2, A Panorama of the Philosophy of Risk](#), in this handbook). A type 1 error or false positive occurs if one establishes an effect (risk) where there is actually none; a type 2 error or false negative occurs if one does not establish an effect (risk) while there is actually an effect. Science traditionally focuses on avoiding type 1 errors to avoid assuming too easily that a certain proposition or hypothesis is true. This methodological choice seems perfectly sound as long as the goal of science is to add to the stock of knowledge, but in contexts in which science is used for practical purposes, as in the case of risk assessment, the choice may be problematic. From a practical or moral point of view it may be worse not to establish a risk while there is one than to wrongly assume a risk. As Cranor (1993) has pointed out the 95% rule for accepting statistical evidence in science is also based on the assumption that type 1 errors are worse than type 2 errors. Rather than simply applying the 95% rule, risk assessors might better try to reduce type 2 errors or balance type 1 against type 2 errors (Cranor 1993, pp. 32–29; Douglas 2009, pp. 104–106).

There are also other methodological decisions and assumptions that impact on the outcomes of risk assessment and the possibilities of error. One example is the extrapolation of

empirically found dose–effect relations of potentially harmful substances to low doses. Often, no empirical data are available for low doses; therefore the found empirical data has to be extrapolated to the low dose region on the basis of certain assumptions. It might, for example, be assumed that the relation between dose and response is linear in the low dose region, but it is also sometimes supposed that substances have a no effect level, that is, that below a certain threshold dose there is no effect. Such methodological decisions can have a huge impact on what risks are considered acceptable. An example concerns the risks of dioxin. On basis of the same empirical data, but employing different assumptions about the relation between dose and response in the low dose region, Canadian and US authorities came to norms for acceptable levels of dioxin exposure to humans that are different by a factor of 1,000 (Covello and Merkhofer 1993, pp. 177–178).

While it is clear that in risk assessment, a number of value judgments and morally relevant methodological judgments need to be made, the implications for the responsibility of risk assessors, as scientists, are less obvious. One possibility would be to consider such choices to be entirely the responsibility of the risk assessors. This, however, does not seem like a very desirable option; although risk assessors without doubt bear some responsibility, it might be better to involve other groups as well, especially those responsible for risk management, in the value judgments to be made. The other extreme would be to restore the value-free science idea as much as possible. Risk assessors might, for example, pass on the scientific results including assumptions they made and related uncertainties. They might even present different results given different assumptions or different scenarios. While it might be a good idea to allow for different interpretations of scientific results, simply passing on all evidence to risk managers, who then can make up their mind does not seem desirable. Such evidence would probably be quite hard if not impossible to understand for risk managers. Scientists have a proper role to play in the interpretation of scientific data, albeit to avoid that data is deliberately wrongly interpreted for political reasons. Hence, rather than endorsing one of those two extremes, one should opt for a joint responsibility of risk assessors and risk managers for making the relevant value judgments while at the same recognizing their specific and different responsibilities. Among others, this would imply recognizing that risk assessment is a process that involves scientific analysis and deliberation (Stern and Feinberg 1996; Douglas 2009).

## Individual Versus Collective Responsibility for Risks

When you get into your car in order to transport your children to school and yourself to an important work meeting, you expose a number of people to the risk of being injured or even killed in an accident. First, you expose yourself to that risk. Second, you expose your children to that risk. Third, you expose other drivers, passengers, pedestrians, and cyclists to that risk. Furthermore, someone made decisions that affected your driving: decisions about driving licenses, street lighting, traffic lights and signs, intersections, roundabouts, and so forth. Who is responsible for these different forms of risk exposure? There is an individual and a collective level at which to answer this question. The underlying philosophical question is that of individual and collective responsibility – to what extent and for which risks is an individual responsible and to what extent and for which risks is society collectively responsible? In the following, we will explain how these issues relate to each other. The analysis of road traffic

serves as an example of how aspects of individual and collective responsibility reoccur in most areas of risk management and policy today.

The fact that you expose yourself to the risks associated with driving a car appears to be a primarily individual responsibility. As a driver with a license you are supposed to know what the relevant risks are. Unless you acted under compulsion or ignorance you are held responsible for your actions, in road traffic as elsewhere. As discussed in section [● Conceptions of Responsibility](#), the condition of voluntariness has been discussed by philosophers since Aristotle. When you voluntarily enter your car and know that you risk yours and others' health and life by driving your car, even if those risks are considered fairly small in probability terms, you are responsible in case something bad happens because you accepted the risks associated with driving. This assessment is, of course, complicated by the behavior of other road users. Perhaps someone else made a mistake or even did something intentionally wrong, thereby causing an accident. In that case, you are often considered responsible to some extent, because you were aware of the risks associated with driving and these risks include being exposed to other people's intentional and unintentional bad behavior. However, other road users may bear the greatest share of responsibility in case their part in the causal chain is greater and their wrongdoing is considered more serious. The point is that the individual perspective distributes responsibility between the individuals involved in the causal chain. The key elements are (1) individuals, (2) causation, and (3) wrongdoing. The one/s that caused the accident by doing something wrong is/are responsible for it. (In section [● Conceptions of Responsibility](#), we mentioned two further conditions for responsibility, i.e., freedom and knowledge. These are usually met in traffic accidents and therefore we do not mention them separately here, but they may be relevant in specific cases.) When attributing responsibility according to this approach the road transport system is taken for granted the way it is. However, as we noted, someone made decisions concerning the road transport system and the way you and your fellow road users are affected by those decisions.

The collective or systemic perspective, instead, focuses on the road transport system. Were the roads of a reasonable standard, was there enough street lighting, and was the speed limit reasonable in relation to the condition and circumstances of the road? The default is to look at what the individuals did and did not do and to take the road transport system as a given and this is often reflected in law. However, in some countries the policy is changing and moving toward a collective or systemic perspective. In 1997, the Swedish government made a decision which has influenced discussions and policies in other European countries. The so-called Vision Zero was adopted, according to which the ultimate goal of traffic safety policy is that no one is killed or seriously injured in road traffic (Nihlén Fahlquist 2006). This may be seen as obvious to many people, but can be contrasted to the cost-benefit approach according to which the benefit of a certain method should always be seen in relation to its cost. Instead of accepting a certain number of fatalities, it was now stated that it is not ethically justifiable to say that 300 or 200 are acceptable numbers of fatalities. In addition to this idea, a new view of responsibility was introduced. According to that approach, individuals are responsible for their road behavior, but the system designers are *ultimately* responsible for traffic safety. This policy decision reflected a change in perspective moving from individual responsibility to collective responsibility. Road traffic should no longer be seen purely as a matter of individual responsibility, but instead the designers of the system (road managers, maintainers, and the automotive industry) have a great role to play and a great share of responsibility for making the roads safer and saving lives in traffic. Instead of merely focusing on individuals, causation, and

wrongdoing, the focus should be on (1) collective actors with the (2) resources and abilities to affect the situation in a positive direction. The example of road traffic illustrates how an activity often has a collective as well as an individual dimension. The adoption of Vision Zero shows that our views on who is responsible for a risky activity, with individual and collective dimensions, can be changed.

Furthermore, this example illuminates the difference between (1) backward-looking and (2) forward-looking responsibility. Sometimes when discussing responsibility, we may refer to the need for someone to give an account for what happened or we blame someone for what happened. In other situations we refer to the aim to appoint someone to solve a problem, the need for someone to act responsibly or to see to it that certain results are achieved. There are several distinctions to be made within these two broad categories, but it could be useful to make this broad distinction between backward-looking and forward-looking notions of responsibility (see also section [❸ Conceptions of Responsibility](#)).

The issue of collective responsibility is a much discussed topic in contemporary philosophy. Some scholars argue that there is no such thing and that only individuals can rightly be considered responsible. This position was taken, for example, by H.D. Lewis (1948) some years after World War II and it is understandable that many people were skeptical to the idea of collective actors and collective guilt at that point in time. The world has changed a lot since then and 65 years after World War II the ideas of collective actors and holding collectives responsible are not as terrifying. On the contrary, against the background of multinational corporations, for example, banks and oil producers, behaving badly and causing harm to individuals it appears more and more crucial to find a way of holding such actors accountable for harm caused by them. Philosophers like Peter French have therefore defended the idea that collective agents, such as corporations or governments, can be morally responsible (e.g., French 1984). Some authors claim that collective responsibility is sometimes irreducible to individual responsibility, that is, a collective can be responsible without any of its members being responsible (French 1984; Gilbert 1989; Pettit 2007; Copp 2007). Others claim that collective responsibility is, in the end, analyzable only in terms of individual responsibility (Miller 2010). The collective responsibility of the government might, for example, be understood as the joint responsibility of the prime minister (as prime minister), other members of the government, members of the Parliament, and maybe civil servants. In section [❸ Further Research: Organizing Responsibility for Risks](#), we will explore possible tensions between individual and collective responsibility, and the so-called problem of many hands.

Scholars are likely to continue discussing whether the notion of collective responsibility makes philosophical sense and if so how it should be conceived. What cannot be denied is that in society we treat some risks as an individual responsibility and others as a collective responsibility. Whereas the risks associated with mountaineering are usually seen as individual responsibility, the risks stemming from nuclear power are seen as collective. However, it is arguably not always that simple to decide whether an individual or a collective is responsible for a certain risk. It is often the case that there is an individual as well as a collective dimension to risks. Climate risks arising from the emissions of carbon dioxide are good examples of this. Arguably, individuals have a responsibility to do what they can to contribute to the reduction of emissions, but governmental and international action is also crucial. Furthermore, it is also a matter of which notion of responsibility we apply to a specific context. While we sometimes blame individuals for having smoked for 40 years thereby causing their own lung cancer, we may make it a collective responsibility to give them proper care.

There are two general perspectives on the balance between individual and collective responsibility for health risks. First, the libertarian approach views lifestyle risks, for example, smoking, as an individual matter and relates causation to blame and responsibility for the cost of damage. A liberal welfare approach considers causation as one thing and paying for the consequences as another thing so that even if an individual is seen as having caused her own lung cancer, she should perhaps not have to pay for the health care she now needs. Furthermore, according to liberal welfare theories, individuals are always situated in a socioeconomic context and, consequently, the fact that a particular individual smokes may not entirely be a matter of free choice. Instead, it may be partly due to the situation she is in, her socioeconomic context, education, and so forth, which entails a different perspective on causation, and hence also on the distribution of responsibility between the individual and the collective. The liberal welfare approach does not pay as much attention to free choice as the libertarian approach, or alternatively does not see choices as free in the same sense as libertarians do. This is because the two perspectives assume different conceptions of liberty. Libertarians focus on so-called negative freedom, that is, being free to do whatever one wants to do as long as one does not infringe on another person's rights. Liberal welfare proponents focus on positive liberty, that is, freedom to act in certain ways and having possibilities to act. The former requires legislation to protect individuals' rights and the latter requires a more expansive institutional setting and taxation to create the circumstances and capabilities (see [Chap. 39, The Capability Approach in Risk Analysis](#), in this handbook) needed for people to make use of those possibilities. Different conceptions of liberty entail different conceptions of responsibility. Those emphasizing negative liberty attribute a greater share of responsibility to individuals and those who prefer positive liberty make governments and societies collectively responsible to a greater extent. (For a classic explanation of the concepts of negative and positive liberty see Isaiah Berlin [1958](#).)

The decision to view a certain risk as an individual or a collective matter entails different strategies for dealing with risk reduction and different strategies for deciding about the acceptability of a risk. If the risk is seen as an individual matter the strategy is likely to emphasize information campaigns at the most. If, for example, road safety is seen as an individual responsibility risk managers who want to reduce the number of fatalities and injuries will inform the public about risky behavior and how to avoid such behavior. "Don't drink and drive" campaigns is an example of that strategy. Some libertarians would probably argue that even this kind of campaign is unacceptable use of taxpayers' money and that an information campaign should only objectively inform about the risks of drunk driving and not give any advice because individuals should be considered competent enough to make their own decisions about driving. However, a "Don't drink and drive" campaign could also be seen as a way of making sure individuals do not harm each other, that is, do not infringe on other individuals' rights not to be harmed, and for this reason it would probably be acceptable to a moderate libertarian. Surely, libertarians would not agree to anything more intrusive than this, for example, surveillance cameras.

If, instead, road safety is seen as a collective responsibility, risk managers may try to find other ways of reducing the risks of driving. In the case of drunk driving, one such example could be alcohol interlocks, that is, a new technology which makes it impossible to drive under the influence of alcohol. This device measures the driver's blood alcohol concentration (BAC) before the car starts, for example, through an exhalation sample, and because it is connected to the car's ignition it will not start if the measured concentration is above the maximum set. Alcohol interlocks are currently used in some vehicles and some contexts in Sweden and elsewhere. It is

possible that the device will be a natural part of all motor vehicles in the future and this would indeed be a way of making drunk driving a collective responsibility, although individuals would still be responsible for not misleading or otherwise circumventing the system.

The collective approach to responsibility for risks is sometimes criticized for being paternalistic. The argument is that people should be free to make their own decisions about which risks are worth taking. One way to assure freedom of choice is to apply the principle of informed consent to decisions about acceptable risk. Informed consent is a principle commonly used in medical experiments and the idea is that those who take part in the experiments are informed about the risks and then decide whether to consent through signing a document. Similarly, individuals are to decide what technological risks they want to take. To this end, they should be informed about the risks of different technologies, and they should be free to decide whether to take a certain risk or not. The approach of informed consent clearly fits in a libertarian approach to risk taking. However, when people make decisions about risks, their choices can be affected through the way information is presented. Thaler and Sunstein (2008) argue that a decision is always made in a context and that “choice architects” design this context. Since choices are always framed in one way or another, you might as well opt for “nudging” people in the “better,” healthier for instance, direction. One example of this is a school cafeteria in which different food products are arranged in one way or another and without removing the less healthy options, a “choice architect” could nudge children in the direction of the healthier options. Even a very anti-paternalistic libertarian, they argue, could accept this since no options are removed and the food has to be arranged in one way or another (Thaler and Sunstein 2008).

There are, however, several problems with applying the principle of informed consent to risk taking. One problem is that it might be hard, if not impossible, to present risks in a neutral and objective way (see also section [Risk Communication](#)). Second, risks are sometimes uncertain. Imagine there is research on how radiation from mobile phones affects grown-ups in the time frame of 10 years after you start using the phone, but not how it affects children or how it affects grown-ups in the long-term perspective of say 20–30 years. When you use your mobile phone or you let your child use one and you have been informed about the known risks, have you consented to all risks of radiation stemming from mobile phones? Third, it might be doubted whether all risks are or can be taken voluntarily, take for example, the risks associated with driving in an area lacking public transport. Fourth, in many cases the decision whether to accept or take a certain risk is or cannot be an individual decision because it affects other people. Take for example, the decision whether a certain area of the Netherlands should be additionally protected against the sea given expectations of rising sea levels due to the greenhouse effect. Such measures are likely to be very costly and whereas some individuals will judge that an increased risk should be accepted rather than spending large amounts of public funds on higher dikes, others are likely to make the opposite assessment.

Decisions about which risks of flooding should be accepted are by their very nature collective decisions. Since such collective decisions are usually based on majority decision making, individual informed consent is not guaranteed. An alternative would be to require consensus, to safeguard informed consent, but that would very likely result in a stalemate and in a perseverance of the status quo. That would in turn lead to the ethical issue of how the status quo is to be understood. For example, in the case of increased likelihood of flooding the question is whether the status quo should be understood in terms of the current risk of flooding, so that maintaining the status quo would mean heightening the dikes, or whether it should be understood in terms of the current height of the dikes and accepting a higher risk of flooding.

Many acts of seemingly individual risk taking have a collective element. Even committing suicide by driving or jumping in front of a train is not an individual act since other road users may come in the way and get hurt and there is probably psychological damage to the train driver and others who see it happen. By driving your car you inevitably risk the lives of others when you risk your own life. Your own risk taking is then intertwined with the risk exposure of others.

The upshot of the above discussion is not that all decisions about risk are or should be, at least partially, collective decisions, but rather that we should distinguish different kinds of risks, some more individual and others more collective. Consider, for example, the alleged health risks of radiation from mobile phones. The risk that is generated by using a mobile phone, and thereby exposing oneself to radiation, is an individual risk; the radiation only affects the user of the phone. Radiation from base stations, on the other hand, is a collective risk. This is why it has been suggested that the former is managed through informed consent whereas the latter should be subject to public participation and democratic decision making (IEGMP 2000).

However, even if we decide that the risks associated with using a mobile phone is sometimes an individual responsibility, it should be noted that a seemingly individual risk carries with it aspects of collective decision making and responsibility since the government and international agencies may have to set a minimal risk level (MRL) stating what is acceptable radiation and what is not. Many contemporary risks are complex and collective. As democratic societies we have to make choices about what risks to allow. There is a procedural dimension to this, but also a normative dimension. As noted by Ferretti (2010), scholars have been discussing how to make sure that the procedure by which decisions about risks are made become more democratic and fair, but that we also have to discuss the normative and substantive issues of what risks are acceptable and what the decisions are about.

## Risk Communication

As we have seen the tasks of risk assessment, risk management and risk reduction involve different groups, such as engineers, scientists, the government, company managers, and the public, with different responsibilities. Since each group has its specific expertise and fulfilling one's specific responsibility often requires information from others, communication between the groups is of essential importance. Risk communication is therefore crucial for the entire system of dealing with risks in order to work.

In the literature, risk communication is often understood as communication between the government and the public (e.g., Covello et al. 1989) (see [Chap. 29, Tools for Risk Communication](#) and [Chap. 27, Emotion, Warnings, and the Ethics of Risk Communication](#), in this handbook). Although as indicated it might be advisable to understand the notion of risk communication broader, we will here follow this convention and understand risk communication as the communication between the government (or a company) and the public. The goals of such risk communication depend to an important extent on whether one conceives of risk management as an individual or collective responsibility as discussed in the previous section. As we saw there, whether risk management is seen as an individual or collective responsibility partly depends on one's philosophical or political stance. However, it also depends on the kind of risks focused on. Moreover, as we argued, risks are often both an individual and collective responsibility. Therefore, the distinction between individual and

collective responsibility does not exactly match comparable distinctions between consequentialist and deontological approaches or between liberal and paternalistic approaches.

If one conceives of risk management, and especially of decisions about acceptable risk, as the individual responsibility of the one taking or undergoing the risk, the responsibility of the government as risk communicator is to inform the public as completely and as accurately as possible. However, it seems that the government should refrain from attempts to convince the public of the seriousness or acceptability of risks. In this frame, the goal of risk communication is to enable informed consent and the responsibility of the risk communicator is basically to provide reliable and relevant information to enable informed consent.

However, if one conceives of decisions about acceptable risk and risk management as a collective responsibility, trying to convince the public of the acceptability or seriousness of certain risks or trying to get their cooperation for certain risk management measures is not necessarily or always morally problematic, especially if the risk communicator is open about his or her goals (cf. Morgan and Lave 1990; Johnson 1999; Thaler and Sunstein 2008). In a liberal society, it might in general be improper for the government to deliberately misinform the public or to enforce certain risk measures, but convincing the public is not necessarily morally problematic. Moreover, in some extreme situations even misinformation and enforcement might be considered acceptable. It is, for example, generally accepted that violence may sometimes be used by the police to reduce the risks of criminality and terrorism. With respect to risk communication, one might wonder whether it would be acceptable to be silent about the risk of burglary if people have to leave their homes as quickly as possible because of the safety risk as a result of a coming hurricane. Misinformation about risks may in some cases be deemed acceptable if the consequences, or risks, of proper information are larger than the risks communicated. In such cases, consequentialist considerations may be considered more relevant than deontological considerations. In general, if one conceives of risk management as collective rather than as a purely individual responsibility, the consequences of risk communication may be relevant to the responsibilities of the risk communicator and these responsibilities can thus extend beyond informing in the public as well as possible. However, it seems that if one accepts that some risk management decisions are a collective responsibility, one can still either take a more consequentialist or a more deontological view on risk communication.

It might seem that the question concerning what information to provide to the public only arises if risk management is (partly) seen as a collective responsibility. If risk management is an individual responsibility and the aim of risk communication is to enable informed consent, the risk communicator should simply pass on all information to the public. However, not all information is equally relevant for informed consent, and so a certain choice of filtering of information seems appropriate. In addition to the question of what information should be provided, ethical questions may arise in relation to the question of how the information is to be framed (Jungermann 1996).

Tversky and Kahneman (1981) have famously shown that the same statistical information framed differently leads to contradictory decisions about what risks are accepted, for example, depending on whether risk information is framed in terms of survival or death. There are many other factors that are relevant for how risks are presented. One issue is the risk measure used. It makes a difference whether you express the maximum dosage of dioxin per day in picograms, milligrams, or kilograms. The latter presentation – maybe unintentionally – gives the impression that the risk is far smaller than in the first case. Another important issue in risk

communication is how uncertainty should be dealt with. Should the risk communicator just communicate the outcome of a risk assessment or also include uncertainty margins? Should the risk communicator explain how the risk assessment was carried out, so that people can check how reliable it is? Should the methodological assumptions and choices made in the risk assessment (section [Conceptions of Responsibility](#)) be explained to the public?

## Further Research: Organizing Responsibility for Risks

---

A major philosophical challenge today is to conceptualize responsibility in relation to collective agency. While the increased control over the environment seems to increase the total amount of responsibility, this responsibility is also increasingly dispersed over many different individuals and organizations. The somewhat paradoxical result is that it sometimes appears to be increasingly difficult to hold someone responsible for certain collective effects like climate change. Partly this may result from the fact that today's society is so obsessed with holding people responsible (blameworthy) that many individuals and organizations try to avoid responsibility rather than to assume it. Ulrich Beck (1992) has described this phenomenon as "organized irresponsibility."

We have identified five important topics for further research that we will discuss in the following subsections. In subsection [The Problem of Many Hands \(PMH\)](#), we discuss what has been called the problem of many hands (PMH).

In subsection [Climate Change as an Example](#), we will discuss the risk of climate change as an example of the PMH. We are all contributing to climate change. However, if and how this observation of (marginal) causal responsibility has implications for moral responsibility is not at all clear and this issue needs considerable attention.

Subsection [Responsibility as a Virtue](#) will discuss the idea that rather than understanding responsibility in a formal way, we should appeal to individuals who should take up responsibility proactively. To that purpose, we suggest to turn to virtue ethics and care ethics. We explore the possibilities of an account of responsibility as the virtue of care, as a way to deal with the PMH.

Another example is the discussion in section [Responsibility for Risks](#) about the related responsibilities for risk assessment, risk management, risk reduction, and risk communication. We have seen that there are some problems with the traditional allocation of responsibilities in which, for example, scientists are only responsible for risk assessment and have no role to play in risk management. These examples illustrate the need to discuss the distribution of responsibility (subsection [The Procedure of Responsibility Distribution](#)) among the actors involved as well as the question of who is responsible for the entire system, which involves the notion of institutional design (subsection [Institutional Design](#)).

### The Problem of Many Hands (PMH)

---

Although Dennis Thompson (1980) already coined the term "problem of many hands" in 1980, relatively little research has been done in this area. For this reason, we will summarize briefly what already has been done, but large parts of the discussion relate to directions and suggestions for further research.

Thompson describes the PMH as “the difficulty even in principle to identify who is responsible for . . . outcomes” (Thompson 1980, p. 905). Many different individuals act in different ways and the joint effect of those actions is an undesired state-of-affairs X, but none of the individuals (1) directly caused X or (2) wanted or intended X. In such cases, it is either difficult to discern how each actor contributed to X or it is unclear what implications the joint causal responsibility should have for the moral responsibility of the individuals whose combined actions caused X. As we have seen, there is backward-looking and forward-looking responsibility. The PMH can be seen as a problem of forward-looking responsibility, but it has primarily been discussed as a problem of backward-looking responsibility. Typically, the PMH occurs when something has happened and although there may not have been any wrongdoing legally speaking, the public may have a feeling that something has been done for which someone is morally responsible. The question is just who should be considered responsible, since the traditional conditions of responsibility are extremely hard to apply.

Two features of contemporary society make the PMH salient today. First, human activities are to an increasing extent carried out by groups of people instead of by individuals. Second, we are increasingly able to control risks and hazards, which also seem to increase our responsibilities. We will briefly discuss these features in turn.

Traditionally, philosophers theorize about morality in relation to individuals and how they act. However, in contemporary societies, a substantial part of the daily lives of individuals are intertwined with collective entities like the state, multinational corporations, nongovernmental organizations, and voluntary associations. Collective agency has become frequent. We talk about nations going to war, companies drilling for oil, governments deciding to build a new hospital, a local Lions club organizing a book fair. As discussed in section ➤ [Individual Versus Collective Responsibility for Risks](#), the concept of collective moral responsibility is a much debated topic in philosophy. A risky activity can be seen as an individual or a collective responsibility, but most risks have aspects of both.

Collectively caused harm complicates ethical analysis. This is so partly for epistemic reasons, that is, because we do not know how the actions of different individuals combine to cause bad things. Furthermore, did each and every individual in that particular collective know what they took part in? However, it is not merely for epistemic reasons that we have problems ascribing responsibility in such cases. Collective harm may also arise due to a tragedy of the commons (Hardin 1968). In a tragedy of the commons, the commons – a shared resource – are exhausted because for each individual it is rational to use the commons as much as possible without limitation. The aggregate result of these individual rational actions, the exhaustion of the common resource so that no individual can continue to use it, is undesirable and in a sense irrational. Many environmental problems can be understood as a tragedy of the commons. Baylor Johnson (2003) has argued that in a tragedy of the commons individuals are not morally required to restrict their use of the common resource as long as no collective agreement has been reached, hence no individual can properly be held responsible for the exhaustion of the commons (for a contra argument, see Braham and van Hees 2010).

Similarly, Philip Pettit (2007) has argued that sometimes no individual can properly be held morally responsible for undesirable collective outcomes (for support of Pettit’s argument see, e.g., Copp (2007), for criticism, see Braham and van Hees (2010), Hindriks (2009), and Miller (2007)). The type of situations he refers to are known as voting paradoxes or discursive dilemmas. Pettit gives the following example. Suppose that three employees (A, B, and C) of a company need to decide together whether a certain safety device should be installed and

suppose that they agree that this should only be done if (1) there is a serious danger ( $p$ ), (2) the device is effective with respect to the danger ( $q$ ), and (3) the costs are bearable ( $r$ ). If and only if all three conditions are met ( $p \wedge q \wedge r$ ) the device is to be installed implying a pay sacrifice ( $s$ ) for all three employees. Now suppose that the judgments of the three individuals on  $p$ ,  $q$ ,  $r$ , and  $s$  are as indicated in the table below. Also suppose that the collective decision is made by majority decision on the individual issues  $p$ ,  $q$ , and  $r$  and then deducing  $s$  from ( $p \wedge q \wedge r$ ). The result would be that the device is installed and that they all have to accept a pay sacrifice. But who is responsible for this outcome? According to Pettit neither A, B, or C can be properly be held responsible for the decision because each of them believed that the safety device was not worth the pay sacrifice and voted accordingly as can be seen from the table (based on the matrix in Pettit 2007, p. 197). Pettit believes that in cases like this the collective can be held responsible even if no individual can properly be held responsible. Like in the case of the tragedy of the commons, this suggests that the collective agency may make it impossible to hold individuals responsible for collective harmful effects.

In addition to the salience of collective agency today, in today's society negative consequences often result from risk rather than being certain beforehand. Whereas moral theories traditionally deal with situations in which the outcome is knowable and well determined, societies today spend a considerable amount of time and money managing risks, that is, situations in which there is a probability of harm. If it is difficult to decide whether killing is always wrong when done by and to individuals, it is even more difficult to decide whether it is acceptable to expose another human being to the risk of, say 1 in 18,000, of being killed in a road crash. Or, on the societal level, are the risks associated with nuclear power ethically acceptable? Would a difference in probabilities matter to the ethical acceptability and if so, where should the line be drawn between acceptable and unacceptable probability? It is difficult to know how to begin to answer these questions within the traditional ethical frameworks (Hansson 2009). The questions concerning the ethical acceptability of risks clearly have implications for responsibility. If A kills B, A is reasonably held responsible for it and the consequences of that vary according to norms and context. If A exposes B to the risk of 1 in 18,000 of being killed in a road crash, in what way is A responsible for that risk exposure? Interestingly, while it appears more intricate to decide how someone is responsible for exposing another person to the risk of dying than it is to decide whether someone is responsible for killing that person, the very concept of risk appears to imply some sense of responsibility. A risk is often seen as something we can or ought to be able to manage and control (cf. section ② Conceptual Relations Between Risk and Responsibility).

**Table 35.1**

The discursive dilemma (based on Pettit 1997)

	Serious danger? ( $p$ )	Effective measure? ( $q$ )	Bearable costs? ( $r$ )	Pay sacrifice $s$ ( $p \wedge q \wedge r$ )
A	No	Yes	Yes	No
B	Yes	No	Yes	No
C	Yes	Yes	No	No
Majority	Yes	Yes	Yes	[Yes] no

Thus, contemporary society is confronted by more collective agency and possibly more risks. These two features put the so-called problem of many hands (PMH) to the fore. A lot may be at stake: people's lives, the environment, and public health. Furthermore, in addition to cases where the probability is relatively well known, technological research and development entail substantial uncertainty about future hazards about which we do not have any knowledge today.

## Climate Change as an Example

---

Climate change is an illustrative example of a substantial risk (or cluster of risks) for which it is extremely difficult to ascribe and distribute responsibility and which is caused by more or less all human beings, private companies, and governments. It is therefore a possible example of the PMH.

In debates about climate change, various notions of responsibility are at play (cf. section [Conceptions of Responsibility](#)) as is reflected in the different principles of responsibility that have been proposed. First, there is the polluter pays principle (PPP) stating that the polluting actor, that is, the one who caused the pollution, is the actor who ought to pay the cost (United Nations 1992; Caney 2005; Shue 1999). This principle applies a backward-looking notion of responsibility since it focuses on the causal link, but it also associates backward- and forward-looking responsibility in the claim that the one who caused the damage is also the one who should rectify the situation.

Second, there is a principle referred to as common, but differentiated responsibilities (CDR), which states that although all countries share responsibility for climate change, the developed nations have a greater share of responsibility to do something about it (forward-looking responsibility) because their past and current causal contribution is greater (backward-looking responsibility) (United Nations 1998). Thus, both the PPP and the CDR assume that the agent who caused climate change is also the one who is responsible to improve the situation. We often think about responsibility in these terms, but it is possible to conceive of responsibility for climate change differently. The ability to pay principle represents a different approach (Caney 2010). Originally, this principle is associated with a progressive tax system to justify why wealthy people should pay a greater share of their incomes in taxes than poor people in order to maintain a social welfare system. It is possible to design a principle of responsibility for climate change in a similar vein. A central principle in ethics is “ought” implies “can,” essentially meaning that it does not make sense to demand that people do X if they are unable to do X. It has also been argued that sometimes “can” implies “ought” (Garvey 2008). This means that it may be reasonable to attribute a greater share of responsibility for climate change to developed nations not only because they contributed more to the causal chain, but because they simply have more resources to do something about it. This would of course not be reasonable for all risks, but considering the scope and potentially devastating consequences of this particular cluster of risks, it may be a reasonable principle in this case.

We have seen that there are different ways of approaching the distribution of responsibility for climate change between collective actors. In addition, there is the question about how to distribute responsibility between individuals versus collective agents, for example, governments and private companies. To what extent are individuals responsible? Furthermore, in what ways and for which parts are they responsible? Some philosophers argue that individuals are responsible, in the sense of accountability and blameworthiness (backward-looking

responsibility) (e.g., Braham and van Hees 2010). Others argue that individuals are not responsible, but that governments are (e.g., Sinnott-Armstrong 2005), and still others argue that individuals are responsible in a forward-looking way, but that they are not to blame for how climate change and environmental problems came about (e.g., Nihlén Fahlquist 2009).

By talking about risks instead of direct harm, we have changed the perspective of time. A risk that something negative could happen is something for which someone can *take responsibility* and do something about. This is different from cases in which harm has already been done. When the risk has materialized, we want to find someone to blame or give an account of what happened. We need a backward-looking notion since harm will be done and we will want to blame someone to compensate victims. However we also need responsible engineering, research, and risk management, that is, people who act responsibly in order to minimize the risks to society, people, and the environment.

The typical PMH situation occurs when something undesirable has happened as a consequence of collective acting. The PMH can be described by the question: “Who did that?” which is the epistemic problem of knowing who actually did something to cause the undesired event, but the PMH can also point to the normative problem that we cannot find anyone whom it would be fair to hold responsible for the undesired event. The responsibility notion assumed in this question appears to be individualistic and backward looking. Although this notion of responsibility is common and in some ways necessary, there are other notions which may complement it. After all, if we are interested in solving the PMH we probably have to look not only for ways to attribute blame when a risk has materialized, but also for ways in which risks can be reduced or managed in a responsible way to prevent them from materializing. Presumably, what we want is to prevent the PMH from occurring. In the following subsections, we will look into three ways to, if not replace, supplement the “Who did that?”-approach to responsibility: responsibility as a virtue (☞ [Responsibility as a Virtue](#)), responsibility distributions (☞ [The Procedure of Responsibility Distribution](#)), and institutional design (☞ [Institutional Design](#)).

## Responsibility as a Virtue

---

Responsibility is an unusually rich concept. Whereas many notions of responsibility focus on attributing blame for undesired events, there is also a notion that focuses on character traits and personality. To be responsible can be more than having caused X, being blameworthy for causing X, or even having particular obligations to do something about X. Responsibility can also be a virtue and a responsible person can be seen as a virtuous person, that is, having the character traits of a responsible person (see section ☞ [Conceptions of Responsibility](#)). We will now take a closer look at this virtue-ethical notion of responsibility.

By researching, developing, and using technology, opportunities are created. In this process, risks are created as well. In essence, technology creates opportunities and threats. It is, in this sense, a double-edged sword. For example, we want oil for energy, which means that we have to deal with risk of leakage as well as the actual leakage when it happens. Although we live with risks every day, it becomes clear to most people only when the risk actually materializes. Intuitively many people probably think that activities providing us with opportunities, but also risks, imply an increased sense of responsibility and that such activities should be carried out responsibly.

To associate the concept of responsibility with character traits and a “sense of responsibility” means having a closer look at virtue ethics (see ☞ [Chap. 33, Risk and Virtue Ethics](#), in

this handbook). Virtue ethics is often mentioned as the third main branch of ethical theories (next to consequentialism and deontology). Virtue ethicists attempt to find answers to questions of what an agent should do by considering the agent's character and the morally relevant features of the situation (Van Hooft 2006, p. 21). Seeing responsibility as a virtue would entail a focus on how to develop and cultivate people's character with the aim to establish a willingness to actively take responsibility. A willingness to take responsibility involves emotions such as feeling personal involvement, commitment, and not leaving it to others, a feeling that it is up to me and a willingness to sacrifice something (Van Hooft 2006, p. 144, see [Chap. 32, Moral Emotions as Guide to Acceptable Risk](#), in this handbook). It is not the same as a willingness to accept blame for things an agent has done wrong (backward looking) although that may be one part (Van Hooft 2006, p. 141). The main focus is on forward-looking responsibility.

One important aspect of responsibility as a virtue is the recognition that being a responsible person is about carefully balancing different moral demands (Williams 2008, p. 459). Against the background of the different kinds of moral demands human beings face today, it may be difficult to point to one action which is the only right one. Instead a virtuous-responsible person uses her judgment and finds a way to respond and optimize, perhaps, the various demands. Against this background, it could be argued that in order to avoid PMH, we need virtuous-responsible people who use their judgment to form a balanced response to conflicting demands. This could be one way of counteracting the organized irresponsibility of contemporary society. The question is of course how such a society or organization can be achieved. Virtue ethicists discuss upbringing, education, and training as ways of making people more virtuous (Hursthause 2000; Aristotle 2000).

As mentioned in section [The Problem of Many Hands \(PMH\)](#), there are two features of contemporary society which combine to put the PMH to the fore. First, collective agency is increasing. Second, the number of risks has increased, or at least our desire to control risks has grown stronger. A virtue approach to responsibility may counteract the problem of many hands in two ways, both related to the second feature. First, focusing on responsible people may prevent risks from materializing instead of distributing responsibility when it has already materialized (see [Chap. 33, Risk and Virtue Ethics](#)). Responsible people are concerned about risks to human health and the environment because they care and they use their judgment to prevent such risks from materializing. Second, when risks do materialize responsible people will not do everything to avoid being blamed, but will take ownership of what happened and make sure the negative consequences are minimized. Whether the virtue notion of responsibility could also meet the challenge of increasing collective agency is less straightforward. It could be argued that the very tendency to have more collective agency counteracts responsibility as a virtue since people can hide behind collective agents. However, the collectivization could also be seen as making it ever more important to develop a sense of responsibility. Such a development would probably start with moral education and training of young children, something which virtue ethicists often suggest as a way to cultivate virtue. It would also require organizations that foster virtues and a sense of responsibility (see also [Institutional Design](#)).

## The Procedure of Responsibility Distribution

As mentioned earlier, the concept of responsibility is extraordinarily rich and refers to not one but many different notions. In addition to the difference between legal and moral

responsibility, there are many different notions of moral responsibility. It is not surprising that people have different notions in mind and what may appear as conflicts about who is responsible for a certain state-of-affairs may sometimes primarily be misunderstandings due to lack of conceptual clarity. However, it is not merely conceptual lack of clarity which causes differences. People disagree about the normative issues involved, that is, how responsibility *should* be understood and distributed. This is true for people in general and surely holds for professionals as well. According to Doorn (2010, 2011), the prevalence of differences in views on responsibility may cause the PMH. One way of attempting to resolve these differences may be to focus on the procedural setting instead of the substantive conception of responsibility. In order to do this, it is important that we agree that there are disagreements. The solution is not to find and apply the right one, but rather to achieve respect for differences, consensus concerning the procedural setting (this may of course be hard to achieve), and possibly agreement on concrete cases of responsibility distributions.

In political philosophy, John Rawls famously showed that what is needed in pluralist societies is a consensus on the basic structure of society among different religious, moral, and other “comprehensive doctrines” (Rawls 1999 [1971]; Rawls 1993). He argues that we cannot expect that all citizens in a pluralist society agree on politics, but there are some basic principles to which most reasonable people regardless of which doctrine they adhere to would agree, not the least because those very principles would grant them the right to hold those different doctrines. In order for people with different comprehensive doctrines to agree to a basic structure as being fair they justify it through working back and forth between different layers of considerations, that is, their (1) considered moral judgments about particular cases, (2) moral principles, and (3) descriptive and normative background theories. When coherence is achieved between these different layers, we have achieved a wide reflective equilibrium (WRE). In spite of differing judgments on particular cases, different moral principles, and background theories, people can justify the basic structure of society. When many people agree on the basic principles of fairness through different WRE we have an overlapping consensus. Therefore, even if the ways in which we justify it may differ substantially everyone agrees on something, that is, the basic structure of society.

Doorn applies Rawls’ theory to the setting of R&D networks (see also Van de Poel and Zwart 2010). The aim is to develop a model which shows how engineers do not have to agree on a specific conception of responsibility as long as they agree on fair terms of cooperation. R&D networks are non-hierarchical and often lack a clear task distribution, which leaves the question of responsibility open. Doorn shows how a focus on the procedure for responsibility distribution instead of a substantive conception of responsibility makes it possible for engineers to agree on a specific distribution of responsibility. They can agree to it because the procedure was morally justified and fair, even if they disagree about a specific notion of responsibility. This way responsibility is distributed, that is, the PMH is avoided, but the professionals do not have to compromise their different views on responsibility. Without reaching a consensus on a responsibility notion or a responsibility distribution, consensus is reached on a procedure yielding legitimate responsibility distributions. In addition to Rawls’ procedural theory there are others theories, for example, based on deliberative democracy that are set out by authors like Habermas, Cohen, and Elster which can be used in order to help focus on the procedure of distributing responsibility instead of the substantive notions (see, e.g., Habermas 1990; Bohman and Rehg 1997; Elster 1998).

## Institutional Design

---

We have discussed responsibility as a virtue ([❶ Responsibility as a Virtue](#)) and the procedure by which responsibility may be distributed ([❷ The Procedure of Responsibility Distribution](#)) as two ways of counteracting the problem of many hands. We will now look at the importance of institutions. In particular, we will look at what has been called institutional design, the purposeful design of institutions (see, e.g., Weimer 1995). Since institutions generally already exist and cannot be designed from scratch, institutional design usually amounts to modulating or changing existing institutions. Institutional design may contribute to solving the PMH in two different ways: (1) it might create the appropriate institutional environment for people to exercise responsibility-as-virtue and (2) it might help to avoid unintended collective consequences of individual actions. We will discuss both possibilities briefly below.

Institutions may facilitate virtuous or vicious behavior. As argued by Hanna Arendt (1965), Eichmann was an ordinary person who, when he found himself in the context of Nazi-Germany, started to behave like an evil person. Institutions may socialize people into evil doing. Although most cases are not as dramatic and tragic as Eichmann's case, the institutions within and through which we act affect to what extent we act as responsible people. Larry May (1996) has developed a theory of responsibility that connects ideas about responsibility as a virtue to institutions. Institutions can facilitate and encourage or obstruct virtuous behavior. May discusses the ways in which our individual beliefs may change at the group level. The community has an important role in shaping the beliefs of individuals. Relationships between people require a certain collective consciousness with common beliefs. The important point about May's theory is that to foster a sense of responsibility-as-virtue among individuals in a group or organization requires an appropriate institutional environment. As we have argued before ([❶ Responsibility as a Virtue](#)) fostering responsibility-as-virtue may contribute to solving the PMH. The additional point is that doing so not only requires attention for individuals and their education but also for their institutional environment.

The other way that institutional design can contribute to solving the PMH is by devising institutions that minimize unintended collective consequences of individual actions. As we have seen in [❷ The Problem of Many Hands \(PMH\)](#), the PMH partly arises because the actions of individuals may in the aggregate result in consequences that were not intended by any of the individuals. The tragedy of the commons and the discursive dilemma were given as examples of such situations. The phenomenon of unintended consequences is, however, much more general. The sociologist Raymond Boudon (1981) has distinguished between two types of systems of interaction. In what he calls functional systems, the behavior of individuals is constrained by roles. A role is defined by "the group of norms to which the holder of the role is supposed to subscribe" (Boudon 1981, p. 40). In interdependent systems, roles are absent but the actors are dependent on each other for the achievement of their goals. An ideal-typical example of an interdependent system is the free economic market. The tragedy of the commons in its classical form also supposes an interdependent system of interaction and the absence of roles, since the actors are not bound by any institutional norms.

According to Boudon, emergent, that is, collective, aggregate effects are much more common in interdependent systems than in functional systems. Reducing emergent effects can therefore often be achieved by organizing an interdependent system into a functional system. This can be done, for example, by the creation of special roles. The "invention" of the role of the government is one example. In cases of technological risks, one may also think of

such roles as a safety officer or safety department within a company or directorate, or an inspectorate for safety within the government. Another approach might be to introduce more general norms as constraints on action. This is in fact often seen as the appropriate way to avoid a tragedy of the commons. In both cases new role responsibilities are created. Such role responsibilities are obviously organizational in origin, but they may entail genuine moral responsibilities under specific conditions, like, for example, that the role obligations are morally allowed and they contribute to morally relevant issues (see also Miller 2010).

## Conclusion

---

There are many different conceptions of risk and psychologists and philosophers have pointed out the need to include more aspects than probabilities and consequences, or costs and benefits when making decisions about the moral acceptability of risks. It remains to be seen whether these additional considerations also need to be built into the very concept of risk.

In addition there are many different notions of responsibility, and the exact relation between risk and responsibility depends on how exactly both notions are understood. Still in general, both risk and responsibility often refer to undesirable consequences and both seem to presuppose the possibility of some degree of control and of making decisions that make a difference.

We started this chapter by mentioning the Deepwater Horizon oil spill in 2010. People were outraged when it occurred and, it seems, rightly so. It raised the issue of responsibility-as-blameworthiness because it appeared as though there had been wrongdoing involved. However, it also raised the issue of responsibility-as-virtue since a lot of people joined the work to relieve the negative consequences of the oil spill and to demand political action to counteract companies from exploiting nature and human beings.

As this example shows, there is not only backward-looking responsibility for risk but also forward-looking responsibility. We discussed in some details relevant forward-looking responsibilities that might be attributed to engineers, risk assessors, risk communicators, and risk managers. We also discussed that risks may be more or less seen as individual or collective responsibility.

We ended with discussing the problem of many hands (PMH), which seems a possible obstacle to taking responsibility for the risks in our modern society. We also suggested three possible ways for dealing with the PMH: responsibility-as-virtue, a procedural approach to responsibility, and institutional design. What is needed is probably a combination of these three approaches, but the discussion suggests that there is also hope in that people are able to unite and release a collective sense of responsibility.

## References

---

- Arendt H (1965) *Eichmann in Jerusalem: a report on the banality of evil*. Viking, New York
- Aristotle (ed) (2000) *Nicomachean ethics*. Cambridge texts in the history of philosophy. Cambridge University Press, Cambridge
- Asveld L, Roeser S (eds) (2009) *The ethics of technological risk*. Earthscan, London
- Beck U (1992) *Risk society; towards a new modernity*. Sage, London
- Berlin I (1958) *Two concepts of liberty*. Clarendon, Oxford

- Bohman J, Rehg W (1997) Deliberative democracy: essays on reason and politics. MIT Press, Cambridge, MA
- Boudon R (1981) The logic of social action. An introduction to sociological analysis. Routledge & Kegan Paul, London
- Bradbury JA (1989) The policy implications of differing concepts of risk. *Sci Technol Hum Values* 14(4):380–399. doi:10.1177/016224398901400404
- Braham M, van Hees M (2010) An anatomy of moral responsibility, manuscript. Available at <http://www.rug.nl/staff/martin.van.hees/MoralAnatomy.pdf>
- Caney S (2005) Cosmopolitan justice, responsibility, and climate change. *Leiden J Int Law* 18(4):747–775
- Caney S (2010) Climate change and the duties of the advantaged. *Crit Rev Int Soc Polit Philos* 13(1):203–228
- Copp D (2007) The collective moral autonomy thesis. *J Soc Philos* 38(3):369–388. doi:10.1111/j.1467-9833.2007.00386.x
- Covello VT, Merkhofer MW (1993) Risk assessment methods. Approaches for assessing health and environmental risks. Plenum, New York
- Covello VT, McCallum DB, Pavlova MT, Task force on environmental cancer and heart and lung disease (U.S.) (1989) Effective risk communication: the role and responsibility of government and nongovernment organizations, vol 4, Contemporary issues in risk analysis. Plenum, New York
- Cranor CF (1993) Regulating toxic substances. A philosophy of science and the law, Environmental ethics and science policy series. Oxford University Press, New York
- Davis M (1998) Thinking like an engineer. Studies in the ethics of a profession. Oxford University Press, New York
- Davis M (forthcoming) “Ain’t no one here but us social forces”: constructing the professional responsibility of engineers. *Sci Eng Ethics*. doi:10.1007/s11948-010-9225-3
- Doorn N (2010) A Rawlsian approach to distribute responsibilities in networks. *Sci Eng Ethics* 16(2):221–249
- Doorn N (2011) Moral responsibility in R&D networks. A procedural approach to distributing responsibilities. Simon Stevin Series in the Philosophy of Technology, Delft
- Douglas M (1985) Risk acceptability according to the social sciences, vol 11, Social research perspectives: occasional reports on current topics. Russell Sage, New York
- Douglas HE (2009) Science, policy, and the value-free ideal. University of Pittsburgh Press, Pittsburgh
- Douglas M, Wildavsky AB (1982) Risk and culture: an essay on the selection of technical and environmental dangers. University of California Press, Berkeley
- Elster J (1998) Deliberative democracy. Cambridge studies in the theory of democracy. Cambridge University Press, Cambridge
- Ferretti M (2010) Risk and distributive justice: the case of regulating new technologies. *Sci Eng Ethics* 16(3):501–515. doi:10.1007/s11948-009-9172-z
- Fischer JM, Ravizza M (1998) Responsibility and control: a theory of moral responsibility. Cambridge studies in philosophy and law. Cambridge University Press, Cambridge
- Frankfurt H (1969) Alternate possibilities and moral responsibility. *J Philos* 66:829–839
- French PA (1984) Collective and corporate responsibility. Columbia University Press, New York
- Garvey J (2008) The ethics of climate change. Right and wrong in a warming world. Continuum, London
- Giddens A (1999) Risk and responsibility. *Modern Law Rev* 62(1):1–10
- Gilbert M (1989) On social facts. International library of philosophy. Routledge, London
- Habermas J (1990) Moral consciousness and communicative action. Studies in contemporary German social thought. MIT Press, Cambridge, MA
- Hansson SO (2008) Risk. The Stanford encyclopedia of philosophy (Winter 2008 Edition)
- Hansson SO (2009) Risk and safety in technology. In: Meijers A (ed) Handbook of the philosophy of science. Philosophy of technology and engineering sciences, vol 9. Elsevier, Oxford, pp 1069–1102
- Hardin G (1968) The tragedy of the commons. *Science* 162:1243–1248
- Harris CE, Pritchard MS, Rabins MJ (2008) Engineering ethics. Concepts and cases, 4th edn. Wadsworth, Belmont
- Hart HLA (1968) Punishment and responsibility: essays in the philosophy of law. Clarendon, Oxford
- Hindriks F (2009) Corporate responsibility and judgement aggregation. *Econ Philos* 25:161–177
- Hunter TA (1997) Designing to codes and standards. In: Dieter GE, Lampman S (eds) ASM handbook, vol 20, Materials selection and design., pp 66–71
- Hursthouse R (2000) On virtue ethics. Oxford University Press, Oxford
- IEGMP (2000) Mobile phones and health (The Stewart report). Independent Expert Group on Mobile Phones
- International Program on Chemical Safety (2004) IPCS risk assessment terminology, vol 1, Harmonization project document. World Health Organisation, Geneva
- Johnson BB (1999) Ethical issues in risk communication: continuing the discussion. *Risk Anal* 19(3):335–348
- Johnson BL (2003) Ethical obligations in a tragedy of the commons. *Environ Values* 12(3):271–287
- Jungermann H (1996) Ethical dilemmas in risk communication. In: Messick M, Tenbrunsel AE (eds) Codes of conduct. Behavioral research into business ethics. Russell Sage, New York, pp 300–317
- Lewis HD (1948) Collective responsibility. *Philosophy* 24(83):3–18

- Martin MW, Schinzinger R (2005) Ethics in engineering, 4th edn. McGraw-Hill, Boston
- May L (1996) The socially responsive self. The University of Chicago Press, Chicago
- Miller S (2007) Against the collective moral autonomy thesis. *J Soc Philos* 38(3):389–409. doi:10.1111/j.1467-9833.2007.00387.x
- Miller S (2010) The moral foundations of social institutions: a philosophical study. Cambridge University Press, New York
- Morgan MG, Lave L (1990) Ethical considerations in risk communication practice and research. *Risk Anal* 10(3):355–358
- National Research Council (1983) Risk assessment in the federal government: managing the process. National Academy Press, Washington, DC
- Nihlén Fahlquist J (2006) Responsibility ascriptions and vision zero. *Accid Anal Prev* 38:1113–1118
- Nihlén Fahlquist J (2009) Moral responsibility for environmental problems – individual or institutional? *J Agric Environ Ethics* 22(2):109–124. doi:10.1007/s10806-008-9134-5
- NSPE (2007) NSPE code of ethics for engineers. National Society of Professional Engineers, USA, <http://www.nspe.org/Ethics/CodeofEthics/index.html>. Accessed 10 September 2010
- Pettit P (2007) Responsibility incorporated. *Ethics* 117:171–201
- Rawls J (1993) Political liberalism. Columbia University Press, New York
- Rawls J (1999 [1971]) A theory of justice, Revised edn. The Belknap Press of Harvard University Press, Cambridge, MA
- Rayner S (1992) Cultural theory and risk analysis. In: Krinsky S, Golding D (eds) Social theories of risk. Praeger, Westport, pp 83–115
- Renn O (1992) Concepts of risk: a classification. In: Krinsky S, Golding D (eds) Social theories of risk. Praeger, Westport, pp 53–79
- Roeser S (2006) The role of emotions in the moral acceptability of risk. *Saf Sci* 44:689–700
- Roeser S (2007) Ethical intuitions about risks. *Saf Sci Monit* 11(3):1–13
- Shrader-Frechette KS (1991a) Reductionist approaches to risk. In: Mayo DG, Hollander RD (eds) Acceptable evidence; science and values in risk management. Oxford University Press, New York, pp 218–248
- Shrader-Frechette KS (1991b) Risk and rationality. Philosophical foundations for populist reform. University of California Press, Berkeley
- Shue H (1999) Global environment and international inequality. *Int Aff* 75(3):531–545
- Sinnott-Armstrong W (2005) It's not my fault: global warming and individual moral obligations. In: Sinnott-Armstrong W, Howarth RB (eds) Perspectives on climate change science, economics, politics, ethics. Elsevier/JAI, Amsterdam, pp 285–307
- Slovic P (2000) The perception of risk. Earthscan, London
- Stern PC, Feinberg HV (1996) Understanding risk. Informing decisions in a democratic society. National Academy Press, Washington
- Thaler RH, Sunstein CR (2008) Nudge: improving decisions about health, wealth, and happiness. Yale University Press, New Haven
- Thompson DF (1980) Moral responsibility and public officials: the problem of many hands. *Am Polit Sci Rev* 74(4):905–916
- Thompson PB, Dean W (1996) Competing conceptions of risk. *Risk Environ Health Saf* 7:361–384
- Tversky A, Kahneman D (1981) The framing of decisions and the psychology of choice. *Science* 211(4481):453–458. doi:10.1126/science.7455683
- United Nations (1992) Rio declaration on environment and development
- United Nations (1998) Kyoto protocol to the United Nations framework convention on climate change
- Van de Poel I (2011) The relation between forward-looking and backward-looking responsibility. In: Vincent N, Van de Poel I, Van den Hoven J (eds) Moral responsibility. Beyond free will and determinism. Springer, Dordrecht
- Van de Poel I, Royakkers L (2011) Ethics, technology and engineering. Wiley-Blackwell
- Van de Poel I, Van Gorp A (2006) The need for ethical reflection in engineering design: the relevance of type of design and design hierarchy. *Sci Technol Hum Values* 31(3):333–360
- Van de Poel I, Zwart SD (2010) Reflective equilibrium in R&D Networks. *Sci Technol Hum Values* 35(2):174–199
- Van Hooft S (2006) Understanding virtue ethics. Acumen, Chesham
- Wallace RJ (1994) Responsibility and the moral sentiments. Harvard University Press, Cambridge, MA
- Weimer DL (ed) (1995) Institutional design. Kluwer, Boston
- Williams G (2008) Responsibility as a virtue. *Ethical Theory Moral Pract* 11(4):455–470
- Wolff J (2006) Risk, fear, blame, shame and the regulation of public safety. *Econ Philos* 22(03):409–427. doi:10.1017/S0266267106001040



# 36 What Is a Fair Distribution of Risk?

Madeleine Hayenjelm

University College London, London, UK

<i>Introduction</i> .....	910
<i>An Equal Distribution of Risk?</i> .....	911
Distributing Goods Versus Distributing Risks .....	911
The Equality Claim .....	913
Indiscriminate Risk Impositions .....	915
<i>Precaution and the Aggregation Worry</i> .....	917
Preamble: Justification to Each Person and the Aggregation Worry .....	917
A Precautionary Thesis .....	918
Two Policies .....	919
Another Look at the Precautionary Thesis .....	920
<i>Fairness and Precaution in Unequal Distributions of Risk</i> .....	921
Justification for Unequal Distributions .....	922
Compensation for Unequal Distributions .....	927
<i>What Is a Fair Distribution of Risk?</i> .....	928
<i>Further Research</i> .....	929

**Abstract:** What is a fair distribution of risk? This chapter will look into three separate, but related, aspects of fairness in risk distributions. Firstly, I will locate the object of fairness when it comes to risk distribution. In contrast to distributions of goods, which we want to both increase and distribute fairly, risks are something we want to decrease and distribute fairly. The question of fairness in risk distributions is the question of how to combine these two partially conflicting claims; to fairness on the one hand and to risk reduction on the other. Secondly, I will take a closer look at what an equal distribution of chances for harm might be. I will point to the problem that the very idea of distributing probabilities entails. Thirdly, I address the question of when deviations from equal distributions of risks may be justified and how such inequalities can be addressed in a fair way. It will be suggested that the locus of fairness of risk should be sought in two steps: (1) the justification of particular risky activities, where the level of risk and the spread of that risk is taken into account, and (2) that any resulting higher risk for certain groups or individuals should be addressed through a combination of consent, precaution, and compensation, seeking to even out unfair exposure.

## Introduction

---

What is a fair distribution of risk? In political philosophy we have since Rawls' *A Theory of Justice* been accustomed to think of equal distributions as fair distributions, unless competing reasons, such as efficiency, calls for some less than equal distribution to be balanced with the claims of equality. In Rawls' theory this balance is struck in the *difference principle* where all deviations from equal distributions need to be justifiable and beneficial also to the least advantaged (Rawls 2003). What is of importance for the present chapter is the underlying idea that equality is the initial starting point for thinking about justice and fairness. The question is when any deviations from equality can be considered fair. According to Rawls, such deviations are only just if they are to everyone's advantage (Rawls 2003, 54ff).

Does this approach to fairness about goods also apply to the topic of risk? Can we adopt the model of balancing equality with risk mitigation, such that *ceteris paribus* equal distribution is to be preferred but may be deviated from in order to reduce the total level of risk? Can we think of risks in the same way being distributed equally so that no one faces more risk than anyone else (unless there is strong risk reducing reasons to distribute them less equally in the interest of all including those at higher risk)? This chapter will investigate the idea of an equal distribution of risk and the justification for when deviations from this ideal of equality may be legitimate. I will argue that the idea of equal distribution of uncertain harm is more problematic than in the case of distribution of goods. Or rather the ideal of equality cannot be properly applied to estimates of probabilities but ought to focus instead on outcomes and sources of risk.

I shall assume that the moral permissibility of imposing a risk upon a person depends, at least to some degree, *upon how much risk others are subjected to*. If one person is subjected to all the risks but no one else is, we would spontaneously think that this would be unfair (Broome 1990, p. 95). Intuitively, an equal distribution seems the most fair. But it is not clear what this would mean: Equal distribution of outcomes or of chances for outcomes? Equal probability for harm given the same exposure or equal exposure for all?

But how much can the spread of a particular distribution tell us about the fairness of that distribution? Within the contractualist tradition, especially following Scanlon, it would seem

that fairness is determined not so much by equal outcomes but equal concern for *each* person (Scanlon 2000, p. 229). The stress is now on the *reasons* for a particular distribution and the competing claims of affected parties. This suggests a different perspective on what an equal distribution would be, one where equality is not to be sought in outcomes, or in chances for outcomes, but in taking the rights or the interests (not to be harmed) of each into equal consideration. Lenman (2008, p. 106) suggests that we could interpret this requirement for equal concern as precautionary measures directed at each.

We can then assume that an equal distribution of risk would be acceptable to each, unless the cost would be too high in terms of increases in the total level of risk. Furthermore, we can also presume that there may be good reasons to allow for less than equal distributions of risk for the right reasons. We shall address these questions one at a time: First, how far will the idea of an equal distribution take us when it comes to fairness in the distribution of risks? Second, how do we distribute unequal shares of higher risks in a fair way? Only after we have answered the above questions shall we attempt to answer the main question: When is a particular distribution of risk fair?

## An Equal Distribution of Risk?

### Distributing Goods Versus Distributing Risks

It would seem that the simplest way to distribute risks in a permissible way would be to distribute risks equally, so that no one faced a greater risk than anyone else. This ideal has a lot of intuitive appeal borrowed from the egalitarian ideals in the distribution of public goods. If there is something desirable for all, one way to distribute it fairly is to divide it equally among those with some claim to it. The same line of reasoning would apply to undesirables or harms. It would then seem only natural to think of *risk of harm* in that same way.

How does this idea of fairness translate onto the topic of risk? Risk is not a divisible public good that we want for each and everyone to have their fair share of. Rather risk is a measurable unwanted side effect of such goods. The two goals to balance in the case of risk are (1) total reduction of risk and (2) equal distribution of risks. Just as we may accept a distribution of goods that is less than equal if it benefits everyone in the name of efficiency or the total amount of goods, we may want to accept a distribution of risk that is less than equal if it benefits everyone in the name of risk reduction. There is the same kind of *leveling down objections* that could be raised against equal distributions of risks as those raised against equal distributions of good (term coined by Parfit in 1997, but objection discussed both before and after). In the standard version, the leveling down objection points out that if equality were of *intrinsic* value, or the *only* value to take into account, then this would lead to absurd or otherwise unwanted consequences (see e.g., Parfit 2002). Equality could demand of us that in cases where it is not possible to raise everyone to the same level of well-being than those who are better off must level down to those who are worse off. This would improve the situation for no one but only serve to make it worse for some. Typically this argument has been brought forward in support for competing theories of equality such as sufficiency or prioritarianism. Frankfurt (1987), for example, defends the sufficiency view, arguing that what matters in cases of fairness is that some people are poor or badly off in absolute terms, not that some people have more. Thus if everyone had sufficient means for their ends, it would not be a problem that some had more than that. Insisting on

redistribution in the name of equality would simply translate into forcing some to level down to the level of those that worse off.

Applied to the risk context the simplest form of the leveling down objection would run something like this. If equality is of intrinsic value then if there is some activity (benefitting to society at large) that brings about a greater risk for some, then everyone must be exposed to that risk. Or, if there is some protective device that would greatly reduce the risk for most individuals but could not do so for everyone (whether some individuals were immune to it or there were not enough to go around) then no one should profit from such protective devices. In the case of risk, we could easily see absurd consequences if we insisted that for every accepted risk everyone must be equally exposed to a similar risk since this would require introducing new risks where they did not naturally occur in order to even out the burden of risk. In other cases, it would lead us to not to employ the safety devices we had access to if they were too few or in some way insufficient to save all. It is for this reason that this chapter follows Rawls' and others in starting from the premise that equality must be balanced against gains for everyone to some degree. In the case of risk this means that there can be cases where we must allow some to use protection that others do not have access to and must allow for cases where we expose risks to some that others are exempt from. In particular, it may be the case that we can reduce the risk for some to start with, before we are able to reduce that risk for everyone.

All in all, it would seem that we can think of fairness and the distribution of risks in much the same vain as that of fairness and the distribution of goods. Altham, for example, suggests that we may think of the “absence of risk” as a primary good among the others to distribute by means of Rawls’ difference principle (Altham 1983–1984, p. 22). Interestingly, Altham does not develop the idea precisely because he does not find it helpful to determine what level of risk would be acceptable. The problem of fairness of risk can then be summarized as the problem of how to strike a balance between fairness and risk reduction.

Risks are not taken or imposed unless they come as part of some package that is advantageous in some respect. We do not just impose a risk on someone, but rather we initiate an activity for the sake of some good that also poses a risk. Nozick expresses this point very well: “For example, it might be decided that mining or running trains is sufficiently valuable to be allowed, even though each presents risks to the passerby no less than compulsory Russian roulette with one bullet and  $n$  chambers (with  $n$  set appropriately), which is prohibited because it is insufficiently valuable” (Nozick 1974, p. 74). We can then add the additional question of how to balance goods and risks (and the distribution of goods in relation to the distribution of potential harm) to the problem of balancing risk reduction and fairness. Take a case of live organ donors, for example, the benefactor will not be the same as the one suffering the risk of donating an organ. Are such cases of risks, when the person who benefits is a different person than the person who runs the risk, fair?

If we return to the parallel case of distribution of goods, there is a further complicating issue when we move from goods to risks due to the uncertainty element of risk. In the ideal case, where there is enough of a particular good for everyone, the egalitarian ideal would suggest that everyone should have an equal share of that good. In other words, the equal distribution entails equal outcomes. In the case of risk even if risks were distributed perfectly equally such that everyone faced the same probability of harm this would not entail equal outcomes in terms of harm. A distribution of risks is not a distribution of harms but of *chances for harm*. The parallel case would be an equal distribution of chances for receiving goods rather than an equal

distribution of goods. But how do we distribute chances equally and what is it that would make equal chances fair if the outcomes are not?

## The Equality Claim

---

A first natural step would be to presume, unless there is a good reason not to, that if we want to distribute risks fairly then what we ought to strive for is *an equal distribution of risk*. Let us refer to this as the Equality Claim (EQ):

EQ: An equal distribution of risk is *ceteris paribus* better than an unequal distribution of risks, even when the unequal distribution would entail a lower total level of risk.

This principle seems to have what we want – it is normatively attractive in the sense that it is fair by respecting equality (not fair in the Rawlsian sense, as that does not require equality). We have yet to say how to weigh it against risk reduction in a fair way. All this claim states is that risk reduction is not by itself a sufficient reason to override an equal distribution of risk. It thereby contrasts with utilitarian ideas of distribution such as risk-benefit analysis, which is a common approach in risk management, and which is insensitive to how risks are distributed as long as they maximize the total reduction of risk (for objections against this approach cf. Roeser and Asvelt 2009). I will however leave the problem of balancing risk reduction and fairness to the side for now and instead focus on the very idea of an equal distribution of risk.

What is an equal distribution of risk? Part of the problem here lies in the ambiguity in the notion of “risk.” This problem becomes even more accentuated when we think of risks as objects to distribute. Hansson (2004, p. 10) identified no less than five different interpretations of risk: an unwanted event that may or may not occur, the cause of such an event, the probability for such an event, the statistical expected value of such an event, and the fact that a decision is made under conditions of known probabilities [for different possible outcomes that may occur]. It is unclear which of these interpretations we have in mind. We know what an equal distribution of goods, or even harms, would be. But in the case of risk, we are neither distributing goods nor harms but *potential* harms. One way to understand equal risk is then to attempt to distribute the *probabilities* for a particular harm equally such that everyone faces the same probability of harm. The EQ claim is then a statement suggesting that an equal distribution of probabilities for a particular harm is better than an unequal distribution of probabilities for that same harm.

One difference between distributing harms and the probabilities of harm is that equal probabilities do not entail equal outcomes (see Broome 1984 for a discussion about *ex ante* versus *ex post* fairness). If we, borrowing an example from Lenman (2008), impose a risk of death with the probability of 1 in 500,000 onto a population of 20 million, then the most likely outcome would be that 40 out of the population die and the rest go unharmed. The kind of fairness that an equal distribution of risk can achieve is that of a fair lottery, not that of an equal amount of well-being. The underlying moral rationale in support of an equal distribution of risk need thus be another one than the one that leads us to prefer equality in distribution of goods and harms that are certain.

What is it about an equal distribution of the prospects of harm, before the actual harm has occurred, that makes us think of it as fair? For one it rules out unattractive alternatives such as selection on ill-willed grounds or for arbitrary reasons. It lends impartiality to the distribution of risk since no one is deliberately favored before anyone else. From this point of view, an equal

distribution of risk would only be fair to the extent that this was the best way to achieve an unbiased or impartial distribution of risk. This, however, could equally well be achieved by applying a fixed rule in the selection of who would be at greater and who at lower risk as long as that rule is applied equally to all, Broome has argued (Broome 1990–1991). It could be objected that a fixed rule is only unbiased in the sense that the decisions would be the same regardless of who made them and thereby arbitrariness would be avoided (see Pettit 1997, Ch. 2 for discussion about arbitrariness and non-arbitrariness), but not the possibility of biased or unfair rules.

Broome interestingly argues that distributions by fair lotteries can be fair in themselves rather than as a means to achieve impartiality (or to break ties between equally strong competing claims) (see e.g., Timmermann 2004 and Taurek 1977 for discussion). In contrast to distribution by selecting one candidate over another candidate on grounds of the stronger claim, in a lottery everyone with a claim, large or small, to a particular good is given a chance to actually get what they have claims to: those with competing but somewhat lesser claims are not simply overridden but are part of the distribution (Broome 1990–1991, p. 94). When fairness in outcomes cannot be achieved, he argues, lotteries can provide “a sort of partial equality in satisfaction [of claims].” He writes: “Each person can be given a sort of surrogate satisfaction. By holding a lottery, each can be given an equal chance of getting the good. This is not a perfect fairness, but it meets the requirement of fairness to some extent” (Broome 1990–1991). He continues: “It does so, of course, only if giving a person a chance of getting the good counts as a surrogate satisfaction of her claim. This seems plausible to me. After all, if you have a chance of getting the good you may actually get it” (Broome 1990–1991).

This line of argument seems to hold in cases where the distribution is entirely controlled like the reoccurring examples of Russian roulette games so often referred to in the ethics of risk literature as a means to illustrate the probability and uncertainty aspects of risk. In its standard adoption to the risk context it goes roughly like this: A game of Russian roulette is played on another person. All the chambers but one are empty such that the person faces a risk of 1/number of chambers risk of death (see e.g., Nozick 1974, Ch 4, 74ff). This example has then been elaborated in all sorts of forms to fit the particular distribution of probabilities for different examples. A risk of 1 in 500,000 can then be translated into a fictive Russian roulette gun of 500,000 chambers of which 499,999 are empty and 1 is loaded. An equal risk of harm imposed on a population of 20 million would then be to impose a risk such that it resembled 20 million independent guns pointed at each one. This way no one would have a greater or lesser chance to be harmed than anybody else. In such a case, we could say about each and everyone out of the 20 million that they faced a 1 in 500,000 chances of death even if only 40 of them actually came to be shot.

One worry here is that we tend to think of such distributions as fair because we associate the idea of equal probabilities to the fairness of a die. But if a die is fair it is because it is expected to land on each side one sixth of the times tossed if repeated a large enough number of times. In other words, the fairness presumes expected equality of outcomes in the totality of events. The question is then if there is anything such as equal chances for one-off games, that will not be repeated (see Ayer 1972, pp. 27–30, about the difference between *a priori* mathematical judgements of chances and actual games of dice). Take the case of the Russian roulette again. Let us presume that 20 million people are subject to the same number of gunmen playing Russian roulette with a one loaded chamber to 500,000 unloaded ones. We may say that the risk is the same for each one of them. We may also conclude that we know something about how

frequent the risk is (1–500,000). But is this sufficient to make it fair? We do not think of a medicine that has a one in 500,000 chances for lethal side effects as particularly fair. Partly this may be because we suspect that even if the guns are the same and the gunmen the same number that if we think of actual cases and actual risks a million and one little differences may influence who actually comes to harm and who does not. (In the case of the Russian roulette where the loaded chamber is placed and how the chambers are spun before they are fired). As Ayer puts it: “If we are going to apply the calculus of chances to actual games, we have to make the assumption that all the logically equal possibilities are equal in fact, and this of course is not a mathematical truth. It is an assumption to which one has to give an empirical meaning” (Ayer 1972, p. 30).

It can be argued that what makes the “equal chances” fair is not some idea of equal shares, but as a means to achieve *impartiality*, to avoid biased or non-neutral decisions (Broome 1990–1991; Mirrlees 1982). In the cases of low probability risks, large populations lotteries would most likely result in a very few people “winning” and most likely not those with the strongest claims and reasons (see Hooker 2005, p. 349 for this point). The conflict here is when random selection is better because more neutral than selection on reasons because they may be biases, or alternatively as tiebreaker when no conclusion can be drawn about competing claims.

Take Broome’s claim about lotteries being fair because everyone has a chance of actually getting the claim satisfied. The lottery metaphor is problematic as soon as we step outside the sphere of abstract fictive cases where we control the probabilities to each person. In most cases, the probability of a particular risk is not about how likely a particular individual is to get this or that but what the frequency of such gains typically is for any member in a particular reference class (Perry 2007). In such a case, a risk may be distributed such that it is true about the activity that it typically results in one death for every 500,000 exposed to it, but that it is not true that each and every person exposed to the risk has a 1/500,000 chance of death. Even if everyone is exposed to the same it may be that due to some medical conditions and other contributing factors only 500 are at genuine risk and the rest at no risk at all. I will refer to such distributions of risk as *indiscriminate distributions of risk*. It seems that in most cases, leaving the hypothetical lotteries and Russian roulette guns aside, that when we talk of equal distributions of risk we refer to indiscriminate distributions of risk rather than equal distributions of probabilities. If this is correct then Broome’s argument that lotteries are fair because everyone has a real chance of actually having their claims satisfied since “if you have a chance of getting the good you may actually get it” simply does not hold.

Thus, to the extent that a risk imposed can be understood as a lottery ticket or a die that gives everyone equal odds for being harmed, the Equality Claim holds (given the *ceteris paribus* before it is weighed against precautionary issues). The second question we need to address is to what extent risks imposed *can* be said to resemble lotteries, Russian roulettes, and games of dice.

## Indiscriminate Risk Impositions

In most cases of risk imposition, what we are distributing is not chances of harm but activities with a certain expected probability of harm. The probability measure is then not describing the risk for any one group of individuals and how likely they are to come to harm but is an estimate of how often such incidents tend to happen for that particular activity. Thus imposing

a risk of 1 in 500,000 on a group of individuals may not imply that any one of them or the whole of the group faces a probability of 1 in 500,000 of harm but that they will be affected by an activity that can be expected to bring about harm in 1 out of 500,000 cases given past occurrences.

We allow chemical emissions to be distributed through factory chimneys, additives into food produce, potentially harmful components in prescriptive drugs, pedestrian crossings over roads but not highways, etc. Take the case of emissions or drugs, how do we know when such distributions are fair? On the one hand we could claim that the factory chimney distributes the risks equally since it just lets it out into the air that everyone in its vicinity breathes. On the other hand, who actually comes to harm may depend upon such factors as how long they are exposed, their medical conditions such as asthma, and whether they are children. We can then say that the source of harm is distributed equally but not the actual risk or the probability of harm, which makes it an indiscriminate distribution of risk, without being an equal distribution of probability of harm. The chances of harm may in such cases be very unequal due to a whole set of circumstances.

This goes even for such cases when we can assess the probability fairly accurately of anyone person out of the population coming to harm. Thus, even if we know for a given harm that the chances are 1 in 500,000 for death, this tells us very little about what determines or influences the unlucky cases. We may distribute a source of harm with the expected risk of death as 1 in 500,000 onto a population of 500,000 and implement that source equally, such that everyone is equally exposed to it, without knowing whether that risk is equally distributed.

Suppose that the risk for a certain population given previous statistics of being hit by lightening each year is  $n$ . However, let us presume that most people who are hit by lightening are in tall buildings or under tall constructions or in some other situation putting them at more risk. Thus not everyone is at the same risk but rather, some are at much greater risk of being struck by lightening than others.

In fact, if the expected number of death is 1 out of 500,000 we may have good grounds to suspect that there are more factors to take into account and that most likely some are at much greater risk than others. The point is that we cannot say what the probability of harm for each individual is and then make sure that everyone has the same probability of harm. Not just because all the causal factors are not known, but because probabilities are estimates of the likelihood of a certain event, typically based on the frequency of it occurring for similar enough cases. A probability estimate is in that sense a qualified guess of how likely something is given what we know about how often a certain harm occurs (for a given reference class). Given this, it is hard to see how probabilities are the kind of thing that can equally be distributed among individuals since they are estimates based on aggregation over many instances and individuals. What we can do is act in response to what we know about possible reference classes with different probability estimates to come to harm from that same source. If we know that some are at greater risk than others, we may seek to lower the risk for that group, or “transfer” them to a lower reference class (such as from the class of smokers to the class of nonsmokers).

What speaks in favor of distributing risks indiscriminately is that in many cases we do not know how the risks fall and if nothing else it can be considered *impartial* in its distribution (not in outcomes but in intentions or motives). In cases where we do not know what particular risk any one individual is facing but only an estimate of how likely anyone of them is to come to harm we have in a way come to knowledge’s end. In such cases, distributing risks indiscriminately may well be the best we can do, presuming that we do not know more about potential

factors that influence the outcomes. But is this sufficient to regard a particular distribution as fair in cases where we both know that the risk to the population as a whole is 1 in 500,000 but also have good reason to believe that some within that population may be at much greater risk than others (even if we do not know who they are)?

The problems discussed thus far when it comes to the idea of equal distribution of risk springs from the fact that probabilities are a way to deal with the uncertainty of the future. It is not chances that we distribute but guesses based on large aggregate data. This means that there is too much room for hidden inequalities for “equal probabilities” to be *an ideal* since it can neither guarantee equal outcomes or equal probabilities for individuals but merely is a way to deal with the uncertainty of who will come to harm. We could, however, think of the sources of risks, and the knowledge we have about their comparative riskiness, as a heuristic tool to get a rough idea about how risks and expected harms will fall. We need not talk about distribution of probabilities to see that placing a nuclear power plant in one place rather than another will, in the case of an accident, affect those closer to it more than those farther away. Thus, placing several industrial plants near the same house will presumably increase the risk from emissions as well as from accidents. Thus, we can distribute chances indirectly by distributing what we can portion out: the sources of risks and the precautionary devices to reduce risks. It would seem that instead of thinking of chances as something to distribute, in trying to seek equality and fairness in terms of risk, we would be much more helped if we sought to identify causal factors that increase the chances of harm and address those causal factors.

## Precaution and the Aggregation Worry

---

Instead of thinking of risk as something we can distribute in a way parallel to other goods we can distribute, we can shift focus to *equal treatment* when it comes to the impositions of risks. Lenman (2008) suggests that what we ought to focus on when it comes to distributions of risk is *precautionary measures aimed at each*. Here a different interpretation of EQ suggests itself: risks should be distributed equally not in the sense of equal probabilities but as *equal precautionary concern* for each person. The concern for risk reduction is here paired with fairness in the way of aiming such risk reduction at each person. First a few things need to be said about Lenman’s contractualist approach before we move on to the idea of precaution.

### Preamble: Justification to Each Person and the Aggregation Worry

---

Lenman’s suggestion stems from Scanlon, both in its insistence on moral justification that takes each person into equal account and in Scanlon’s stress on the role of precaution (Scanlon 2000, Ch. 5, esp. 229ff). Lenman argues that contractualism, in particular of Scanlon’s kind, is better suited to take care of our moral intuitions with regard to fair distributions of risks than consequentialism is because of its stress on each individual’s importance in moral justification (Scanlon 2000). As Scanlon puts it: “the justifiability of a moral principle depends only on various individuals’ reasons for objecting to that principle and alternatives to it” (Scanlon 2000, p. 229). The contractualist starting point for Lenman’s investigation is explained thus:

- ▶ The key distinguishing feature of contractualism, for my purposes, is the thought that the right normative ethical claims are those who are best able to justify to others where, crucially, this

justifiability is understood as justifiability to *each other person*. . . . This kind of contractualism contrasts with rival utilitarian views in insisting that we may not simply aggregate costs and benefits to different persons and seek to maximize this aggregate. Harm to you cannot be straightforwardly compensated by benefits to me (Lenman, 99ff, italics in original).

This last point is not unique to contractualism but would hold for any non-consequentialist critique of cost-benefit approaches to risk. The contractualist take is that any distribution of risk must be such that it is acceptable to each person affected (as if they were to sign a contract with each other), the focus being on the role of reasons.

In the context of distribution of risks, a particular distribution can be considered permissible only if it is justifiable to each of the persons affected by that distribution. That the majority benefits from a particular distribution is not sufficient to make it permissible. Only such reasons that those who are getting the lesser deal can also accept can make a particular distribution permissible. If we allow aggregation of benefits and costs across a population, we may come to violate the Kantian requirement that we must not treat others merely as means and that harms to one person cannot be outweighed by the benefits to others. Let us refer to this idea as *the aggregation worry*: an increase in risk for some cannot be justified by aggregating the benefits for others benefitting from that distribution (see Scanlon 2000, pp. 229–241, esp. p. 235; Ashford 2003; Reibetanz 1998; Parfit 2003). This is simply a version of the classical objection to utilitarianism that it matters who benefits and who is harmed and that harm to one person cannot be justified by the benefit to others, regardless of how many those others may be or how great that benefit is. Risks imposed must be such that they are acceptable to each, or in Scanlon's terminology, such that no one reasonably can reject them. Let us therefore specify *the aggregation worry* further: A distribution of risk is not justified merely by the aggregate total benefits, but the risks and benefits must be distributed in a way that is justifiable to all. A fair distribution of risk would then be one that could not be reasonably rejected by anyone and one that treated each person equally in giving each of their interests equal weight.

## A Precautionary Thesis

---

Now let us return to the precautionary aspect, introduced in the beginning of this section, and the idea that risks must be implemented in a way that precautionary measures are directed at each.

Lenman introduces his *precautionary thesis* in two steps. The first step is to merely acknowledge the obligation to reduce risk or to “contain it” as he puts it. The second step is to insist on the need for the justifiability to each person. Thus, in the first instance the precautionary thesis merely suggests: “. . . where risk of harm to people cannot be avoided except at unreasonable cost, we should take reasonable precautions to contain that risk” (Lenman, p. 106). Again, this risk reduction element has not specified any requirements on how such precautions are achieved or how this is done in fair way. Lenman writes:

- ▶ If we are to impose risks on others, each of them might very reasonably require of us as a condition of accepting this, that we take all reasonable precautions to avoid their coming to harm. If you have set aside a special site for the disposal of dangerous waste, you should put a fence around it, post signs warning of the danger, and so forth. If you are going to drive your car,

you should do so carefully, having made reasonable checks that the car is in a safe condition. Such precautions do not mean that nobody will be hurt or killed. Indeed, we can often be pretty sure some people will be hurt and killed as a consequence of the risks we all accept. That reasonable precautions are in place plausibly makes this acceptable (Lenman 2008, p. 106).

The next step is then to stress that precautionary actions must be directed at each individual.

- ▶ We may thus note that for a contractualist it will not suffice to insist that for a population on whom we impose some risk, we must seek to minimize the risk of harm to them by taking precautions. We might insist further that we seek to minimize the risk of harm to *each* of them, aiming any precautions we take at the safety of *each* affected person (Lenman, p. 106).

Thus far we have a normative restriction saying that we must minimize the risk of harm to each individual by taking precautions and that such precautions must be aimed to each person's safety. It is this twofold claim that Lenman refers to as *the precautionary thesis*.

## Two Policies

---

Lenman addresses the topic of precaution and risk distributions by way of a number of hypothetical cases. In each case there is some risk to the general public that can be reduced in various ways. In the example below, there is an initial 1 in 500,000 risks of death imposed on the entire population of 20 million. The choice is now between two possible risk-reducing policies:

- ▶ The first (Policy E) is to reduce the risk to each of the 20 million to 1 in one million.
- ▶ The second (Policy F) is to reduce the risk to 19 million of them to 1 in 19 million, while the risk to the remaining one million is increased to 1 in 100,000 (Lenman 2008, p. 107).

The example is intended to bring out the difference between the contractualist and the consequentialist approach when it comes to risk distributions and precautionary measures directed at each. The crucial point is that Policy F is clearly preferable from a consequentialist point of view since the risk reduction is larger (from 40 to 11 expected number of deaths compared to 20 in Policy E) but achieves this by disregarding the precautionary interests of the one million it increases the risk for. In Policy E, the reduction is the same for everyone. Policy F thus violates the precautionary thesis of aiming precautionary measures to each since instead of lowering it for all individuals a minority has their risk increased with this policy. Increasing a risk for some in order to decrease the total would make it a case of precisely the sort the aggregation worry that Lenman seeks to avoid. Lenman therefore prefers Policy E, “for in E we may be in a position to satisfy each person that we have taken reasonable precaution to avoid their coming in harm’s way. Moreover, we may be able to do this in spite of the fact that we know some of them *will* come in harm’s way” (Lenman 2008, p. 107).

But what is the difference between the two policies? In E there is a general reduction of risk. In F there is a reduction for the majority achieved through an increase for a minority. This, Lenman seems to suggest is sufficient to make distribution like that in F, in most cases, unfair for the minority who had their risk increased. They can claim that precautionary measures have not been directed at them at all since their risk increased. What is interesting here is that Lenman argues that it is *the structure* of the redistribution in F that makes it morally

problematic (Lenman 2008, p. 107). The structure he has in mind is that of imposing a greater risk on some individuals for the sake of risk reduction for the many. Yet, if the very structure of this kind of redistribution would always be a reason counting against it then such reasons must count against all cases where we reduce risk by the aid of someone actively reducing that risk at some increased risk to himself or herself. We would need to exclude cases like reducing the risk of fires by letting fire-fighters combat fires. The reason we may find such risks acceptable may however be due to voluntariness and consent. The point here is directed at Lenman's argument that it is the structure of distributions that ought to worry us. Fairness sought by means of avoiding anyone being put at greater risk seems to come at a high cost in terms of risk reduction since many risks cannot be reduced without putting some at greater risk in order to reduce it. In any society, there is the need for some to take on greater risks for the sake of others and some risks will occur in greater proximity to some individuals than others. Rather than excluding cases like Policy F, we need to ask in which circumstances are such policies acceptable and when not? Here other factors presumably play a role: How high is the higher level of risk? Which precautionary measures are in place in order to lower that higher risk? How justified is the mission? Is there any way to get someone to freely volunteer to take that risk (i.e., not forced by unfortunate circumstances)? What kind of compensation can be expected?

Policy F is underdescribed. It is both compatible with cases like the volunteering, especially trained, and compensated fire fighters as well as cases where a government would deliberately target a group of dissidents to be exposed to a higher risk. Lenman seems to rely on the contrast between E and F to the explanatory work, or, rather, the presumed fairness of E. But given the nature of probabilities, E does not guarantee an equal distribution to each one, but merely a general distribution at a certain level of risk with the actual distribution unknown. The difference then comes down to whether it is worse to distribute a risk at a certain level with the actual risks to each person unknown or to distribute it such that a known group is exposed to a larger risk. In favor of E speaks an argument from intentions: It is morally better to not deliberately aim at increasing the risk for any one group. In favor of F speak arguments from consent, precaution, and compensation: It is only when we know who is at greater risk that we can obtain their consent, tailor precautionary measures to their greater risk, and compensate them *ex ante* for taking such risks.

## Another Look at the Precautionary Thesis

Let us return to the idea that aiming precautionary measures at each person affected as equal treatment and thus as an alternative to equal distribution of chances. Let us first distinguish between on the one hand the distribution of a particular risk (as it is imposed or occurs naturally) and the precautionary activities implemented to protect individuals from that risk once in place. Let us place the aim to minimize the harm to each in the former category and the aiming of precautions at the safety of each affected in the latter category. These two may or may not match each other. The notion of precaution seems to presume some level of risk already expected that the precautions are directed at. Thus, we have two different levels to apply fairness to: the level of justification for a particular imposition of risk and the precautions directed at a person. We could, for example, have dangerous road crossings near some schools but put in extra traffic policemen or volunteers to help children across the roads near the safest roads. There are thus several dimensions to fairness and spread of risk here: we

can have an even or uneven spread of risk; we can have an even or uneven spread of preventive measures, and furthermore those preventive measures can match more or less well those at greater risk in the spread. In order to bring out this distinction let us reformulate the precautionary thesis thus:

Any risk policy must: (a) seek to minimize the risk of harm to each person in the population and (b) aim precautions to each person affected by that risk.

Read in this way, the precautionary claim consists of two rather separate requirements on fairness and risk reduction: on the one hand an obligation to lower the level of risk and on the other hand a requirement to treat each person equally when planning precautionary measures. This would allow us to address both claims to fairness and precautions separately, and furthermore to employ precautionary aims to adjust inequalities that arise through inequalities that would arise on the level of risk reduction or distributions generally. This idea of fairness and risk does not imply equal risk or a particular distribution, only that a particular distribution is only acceptable if it seeks to minimize the total level of risk *and* anyone who is affected by that risk is treated equally with regard to implemented means to counteract, reduce, and protect them from that risk in a way that would not violate the aggregation worry or the stress on reasons acceptable to each.

Lenman, however, does not explore this possibility since he does not acknowledge the difference in seeking to minimize the risk to each and aiming precautionary measures at each. In other words, introducing a risk-reducing policy such as E or F seems to be taken to be the same as to implement precautionary measures in that they aim at risk reduction. Or, put differently, to direct greater risk onto some individuals is incompatible with aiming precautions at that same person.

## Fairness and Precaution in Unequal Distributions of Risk

---

Let us return to the question of what a fair distribution of risk is, as defined by on the one hand the need to reduce risks and on the other to do so fairly. We have discussed the ideal of equality in risk distributions. Let us briefly return to this idea of equal distribution, not of exact probabilities but as a goal to achieve approximate equality in terms of equal concern or respect in risk impositions, or at the very minimum, in planning how to distribute resources of precaution or when imposing higher risks. I addressed this question in terms of the aggregation worry: Risks imposed cannot be justified in terms of the total benefits it brings. How do we then address policies where an individual or a group is put at greater risk than the rest? There is, as I have said before, a conflict between on the one hand fairness and the reduction of risk on the other. In particular, if we take the reduction of risk as an action that needs to be performed by someone or other in order to reduce the total risk and fairness to require that we do not use others merely as means for the benefit of others. On the one hand we have a precautionary thesis that asks us to whenever possible to lower the risk or at the very least keep it at a low level. According to Lenman's precautionary thesis, for example, we have, when "a risk of harm cannot be avoided except at unreasonable cost," an obligation to take "reasonable precautions to contain that risk" (Lenman 2008, p. 106). On the other hand, we have a requirement that tells us not to use others as means and not to increase risk for some in order to lower it in total. A concern is that a too strict application of the aggregation worry may rule out too many cases of risk reduction where someone is called upon to bring about that reduction of risk.

Arguably, most particular risk reduction involves increased expected risks to some of the population.

Now let us turn to two questions: (1) Is it ever justified to impose a greater risk onto a person in order to reduce the general level of risk? (2) If so, how can that justified greater level of risk be duly compensated or in other ways addressed such that it can also be considered fair?

## Justification for Unequal Distributions

---

Before I turn to the question of whether it is ever justified to impose a greater risk onto a person than the level of risk that others are facing, I spell out more clearly ways in which we may impose a higher risk on an individual.

In some cases, I have referred to those as *indiscriminate* impositions of risk, we simply do not have the relevant knowledge to know *that* some will be more at risk than others. We may think that we impose the risk to the same degree to everyone, only that some of those individuals, unknown to us, are more vulnerable to those risks. The permissibility of such cases depends upon what stand we take on moral obligations to know.

In other cases, what is sometimes referred to as cases of *anonymous risks*, we know that imposing a risk onto a large population will lead to some expected number of deaths but *we do not know who* are more likely to die. It may be that we had more knowledge about the differences within that population and about the active causal ways of that risk that we would be able to further narrow down our predictions and identify certain high-risk groups. These kinds of cases can be roughly divided into two categories: those cases where who will come to harm is merely a case of bad luck (a pure Russian roulette case) and those cases where those who will come to harm were more vulnerable in ways that could have been predicted given more knowledge.

Leaving both indiscriminate and anonymous risk impositions to the side we are left with the following two kinds of cases:

1. *Unequal imposing of risk.* We impose a source of risk such that only some are exposed to its hazard and others are not at all exposed to it or at a much smaller level of risk.
2. *Unequal provision of protection.* We provide safety measures to some but not all out of a population exposed to a risk.

In both cases, we impose a risk, or introduce a risk policy, such that we know that this will affect an individual or a group of individuals more than the others. Unequal provision of protection (case 2) for cases of equal exposure of risk would simply be a matter of either scarce resources where the case needs to be solved in an as fair manner as possible (as through a lottery), or it is based on arbitrary or discriminatory reasons. Unequal provision of protection can however be a case of matching unequal distributions of risk (combination of cases 1 and 2) such that those at higher risk are also better protected in a way that reduce their higher risk whether or not it is reduced to the same level of risk as those at lower risk with less protection. I will return to this aspect of unequal provision of protection in the section about compensation below. For now let us turn to the question of knowingly imposing a higher risk onto somebody (case 1).

Some public goods are of the nature that they bring a certain level of risk with them either in their making, in their operating, or by the mere fact that they have to be situated closer to

some individuals than others. When is it justified to impose a greater risk onto some in order to achieve such goods?

There is a distinction that we need to keep in mind made here between the justification for an activity that also happens to be risky and the justification for exposing some individuals to that risk. The justification for building a bridge comes from its value. The justification for exposing the constructors of that bridge comes partly from the value it is expected to bring and partly from how safe it is to build that bridge and how worthwhile it would be for those workers. That everyone can reasonably accept the need for building bridges is not sufficient to conclude that it would be reasonable to put the workers to any level of risk to build it or that it would not matter whom it benefits. We may think that the harm resulting from untamed fires are so devastating that we need to allot the task to put fires out to some individuals even though this would increase the risk to them. Whether the way that is done in a fair way that makes up for the greater risk or not is a separate issue. In other words, even if we accept that some may be put at greater risk than others if the public good is important enough this does not mean that we can require such services at any cost.

I shall assume that any deviations from equal distributions need to address two rather separate lines of justification in order to be morally acceptable. First, if we expect some individuals to be put at higher risk for some public interest this requires a higher demand on the justification for that activity than otherwise. Activities that would have been justified may no longer be so when they come at the cost of asking someone to put himself or herself at risk for it. Second, even if a particular activity were justified even at the cost of some being at a higher risk, this would still require us to address that greater burden for those expected to risk their lives. In particular, I shall argue that in this more specific case the following considerations are of central importance: the degree of consent, the level of risk, and level of compensation. If the second line of justification is not satisfactorily answered, then this may affect the justification of the activity or spread of risk itself (although there may be cases where the activity is of such importance that the activity is justified even when the demands of the second line of justification are not satisfied). The point here is that in order to be justified in imposing a greater risk on some, at the very minimum, we need to make sure that the project that requires that level of risk is worthwhile, presumably in a way that is convincing also for those at risk.

There is a conflict of interests here between those who would benefit from the goods to be achieved by a risky activity and those whose lives would need to be risked for it. Scanlon brings out this conflict in the following way:

- ▶ Suppose, then, that we are considering a principle that allows projects to proceed, even though they involve risk of serious harm to some, provided that a certain level of care has been taken to reduce these risks. It is obvious what the generic reason would be for rejecting such a principle from the standpoint of someone who is seriously injured despite the precautions that have been taken. On the other side, however, those who would benefit, directly or indirectly, from the many activities that the principle would permit may have good generic reason to object to a more stringent requirement. In meeting the level of care demanded by the principle, they might argue, they have done enough to protect others from harm. Refusing to allow activities that meet this level of care, they could claim, impose unacceptable constraint on their lives (Scanlon 2000, 236ff).

The above conflict has parallels to *the problem of paralysis* that in particular rights-based theories of risk would run into. If we were to have a right that others do not impose risks of

harm on us, then their many everyday activities would be ruled out (see Hayenjelm and Wolff, *forthcoming*, for overview). The problem, as pointed out by Nozick (1974, 74ff) and later discussed by others (see e.g., Railton 1983 and Hansson 2003, 297ff) is that traditional rights theory does not come in degrees. Thus if we have a right that others do not impose a risk of harm upon us then it does not matter how small that risk is since the point with rights is that they mark out firm moral boundaries that may not be trespassed unless consented to by the person (Nozick 1974, 74ff; Railton 1983, p. 3). Here, if the claims against risk of harm for an individual were greater than the benefits for another individual, then this would suggest that such activities were not to be permitted no matter how many individuals benefited from that activity since we are not allowed to aggregate interests according to rights-based theories of risk.

The interesting thing to note again is Scanlon's stress on precaution when it comes to the justification of risk. The cases to reasonably reject are those where the reasonable precautions to avoid harm have not been taken. To rule out all cases that would pose a risk of harm would be too restrictive. He writes: "Our idea of 'reasonable precautions' define the level of care we think can be demanded: a principle that demanded more than this would be too confining, and could reasonably be rejected on that ground" (Scanlon 2000, p. 209). Let us take notice of the fact that the justification for unequal risk, on his view, thus depends to a large part on *the level of protection* and care directed to those at greater risk.

Others have stressed the degree to which the benefits in question befall those put at greater risk. Although it seems obviously unfair to let someone run a risk for the benefit of others, it is not clear what the claim to benefit can achieve on its own in terms of justification (Lenman 2008; Hansson 2003). In some cases such as individual risk taking, the very fact that the person running the risk is the person who also is the sole beneficiary of it is what makes it morally relatively unproblematic. But in the cases of large projects that impose a greater risk on some, whatever benefit it will bring to the person at higher risk, is presumably the very same benefit that it will bring to others. Thus, although it would be unfair to exclude the person at risk from the benefits, or to merely make him or her work for the benefit of others, giving him or her an equal share in those benefits cannot justify his or her greater risk. It could be argued that what we need to provide is a larger share of the benefits, but this is hardly likely to be possible in cases of public goods such as the construction of tunnels and bridges. The justification for such cases of increased risk cannot be fulfilled looking at the level of benefits alone.

This problem, however, only arises if we look at the risks and benefits for each case of risk impositions separately. Hansson (2003) suggests a reciprocal system of risks, such that each risks does not need to be beneficial to the person at risk, but that the social system as a whole need to be. He introduces the following example:

- ▶ In your neighbourhood there is a factory that produces product A, which you do not use. The factory emits a chemical substance that gives rise to a very small risk to your health. At the same time, another factory, far away from your home, emits other chemicals in the production of product B that you use. One of the neighbours of this second factory does not use product B, but instead uses product A. In this way, and sometimes in much more complex chains, we may be said to exchange risks and benefits with each other (Hansson 2003, p. 305).

This approach suggests a way forward, perhaps much along the lines of Rawls' ideal of a fair cooperative system. The point Hansson makes is that not each project needs to be beneficial for the person who is at greater risk from that project, if there are other risks that he or she does benefit from in a cooperative system working to everyone's advantage. However, even such

a system that works to everyone's advantage is compatible with unequal distributions of risks and benefits. There may for example, Hansson suggests, be two different groups in society: one that benefits largely and faces smaller risks and another that benefits much less but carries more of the risk burden. Hansson therefore adds an equity requirement: "Exposure of a person to a risk is acceptable if and only if this exposure is part of an equitable social system of risk-taking that works to her advantage" (Hansson 2003, p. 305). Hansson does not develop what, more precisely, an "equitable social system of risk-taking" would entail. This principle seems to fit well for a system of exchange of comparable risks and benefits such as the mentioned example with the two factories. But it is unclear whether it can provide much guidance for odd risks that pose so much greater risks than the average risks within such a system that they are not compensated for. We could, for example, presume a system where all young men were expected to enrol into military service and all women were expected to give birth and that at some point in time these two risks would be comparable in the number of expected deaths. Let us also presume that both these kinds of risks were at this point considered equally beneficial to society. Now, consider a change for the better, that either military service or birthgiving is made much safer, but not both. Presumably the benefits would be the same. In such a case, it is not obvious what Hansson's principle would suggest. Would it make the now less safe risk impermissible? Would it require greater benefits for those taking the greater risks than before? Would it simply require us to reduce the other risk to the same degree as before? Would it require us to protect those at greater risk with greater protection from all other risks? Similar problems arise with all risky activities that are not obviously counterbalanced by other comparable risks or extra benefits. Consider the benefits from information technology benefiting most or all and the extra risks such technologies require from miners to retrieve the necessary minerals and metals for circuit boards and other hardware.

Furthermore, there is the question of risks that affect those that choose to not be part of the social system or do not enjoy the benefits unanswered such as the problem of the risk of falling planes and the Amish farmer who does not enjoy any of the fruits of technology (Munoz-Dardé, forthcoming).

There is a third aspect, besides level of precaution and share in benefits, which affects the justification for activities that put some under greater risk. This aspect is that of voluntariness or consent. This aspect comes most clearly into view when we ask someone to do something for the rest of society that is either not balanced out in a reciprocal way, as participating in a medical trial or taking up a place in the army, or due to some special skill, talent, training, or knowledge that they have that makes them particularly well-suited for this task. Scanlon, for example, already presumes that his higher-risk workers engage in the risky project voluntarily: "Our sense that it is permissible to undertake these projects also depends crucially on the assumption that precautions have been taken to make the work safe and that, in addition, workers have the choice of whether or not to undertake the risks involved" (Scanlon 2000, p. 236).

To some extent the justification for imposing greater risks presumes consent. In some cases, we are not in a position to obtain such consent for practical reasons. Nozick mentions the cases of small or diffuse risks onto a large public where we do not know who will be put at risk (Nozick 1974). Other cases where such consent would not be possible concern deflecting a bomb from a larger city onto the countryside (a much discussed example introduced by Thomson) and examples related to the so-called *doctrine of the double effect*, where a beneficial activity has foreseen but unintended side effects (Thomson 1986, p. 89). In such cases, consent is not possible because there is no time, or we simply do not know who will come to harm.

There is a kind of case that is of particular interest because it brings out the conflict between the risk reduction aim and the fairness aim more clearly. I am here thinking of the cases where there is a public risk to all that can only be lowered if some particular individual or individuals take on a greater burden. Do we have a right to send the trained soldier, the trained policeman, the trained fire fighter onto missions of clear risk, or would it be fairer to conduct a lottery and select anyone out of the population? Consider the following case:

A threat of some sort is putting an entire population of a hundred individuals at risk. Let us say it is a tiger attacking the village (or an invading army or a bush fire). You, being the politically elected head over this village can choose to do one of two things:

- (a) Make the skilled person P, with risk to his own life, confront the threat for the safety of the population at large. If he succeeds the threat is completely abolished (let us say the tiger is slain). The probability that he will succeed is 0.8 and the probability that he will die while trying and failing is 0.2.
- (b) Conduct a lottery where any one out of the population is made to tackle the threat with risk to his, or her, own life. If he or she succeeds the threat is completely abolished. The average probability that any random man or woman succeeds is 0.2, the risk to his or her own life 0.8 (and no one besides P has a greater chance at killing the tiger than 0.3).

The rationale for choosing (a) could easily be motivated from a consequentialist point of view. It would simply minimize harm for all involved parties. But this rationale will not do if we look for a contractualist justification since this would precisely lead us into the lap of the aggregation worry. The lottery in (b) could on the other hand be defended since everyone would be doing their fair share for something that is the common interest of all. It would seem that the contractualist ought to prefer the lottery, even at the foreseeable cost of several people dying trying, and the tiger still at large. But that scenario is clearly suboptimal; every person selected by the lot is put to a much greater risk than necessary, given that P could have done it at a much smaller risk, for the benefit of all. Cases of a similar structure have been addressed in the literature, but often with the assumption that the person put at risk would almost certainly die. Hansson discusses a similar case where a repairman can be sent in to repair a gas leak at a 0.9 risk of death for himself, or this gas can be let out into the environment at the risk of 0.001 for the entire population (Hansson 2003, cross-reference to Hansson's chapter). In such cases, it seems that the risk for the person in question is so high that we could not reasonably require anybody to take on such missions.

Broome discusses a case where one man is more skilled than anybody else to take on a dangerous mission but no more likely to survive than the rest.

“Someone has to be sent on a mission that is so dangerous she will probably be killed. The people available are similar in all respects, except that one has a special talent that make her more likely than others to carry out the mission well (but not more likely to survive). This fact is recognized by her and everyone else” (Broome 1990–1991, p. 90). In Broome’s case the success of the task does not imply a greater chance for survival, and almost certain death is expected. In such cases, Broome suggests that it may be unfair to send the talented person and in fact more fair to conduct a lottery among all of them. If the mission is a threat to all, however, and the chances of failure is in itself a danger it may be the case that we ought to send the talented man. But, Broome argues, this would be a case where other interests are weighed against fairness, not what fairness requires.

Such extreme suicide missions do not answer our more modest question: When is it justified to reduce a risk by expecting someone tackle a risk for everyone else’s sake? On what

grounds can we expect others to fly planes, give birth, fight fires, and fight wars? In many such cases where there is professional expectation to tackle a risk there is also the skills and training such that not only is the chance of success greater but the actual risk is comparatively speaking lower than had anybody else taken on that particular kind of risk. In such cases, conducting a lottery would mean that we would put someone at higher risk than necessary rather than that the person most skilled and who would be at a lower risk. Furthermore, everyone else would be at higher risk since the probability of failure would increase and it could be claimed that the appropriate level of care had not been aimed at them by letting someone unskilled to do the job. What is missing from these examples is the role of consent.

Thus far we can conclude that part of the picture when assessing the justification for activities that require that someone is put at higher risk for the sake of everyone else we need to address all these aspects: the level of risk for the person at higher risk and for the public at large, the benefits of the project in general and for the person at higher risk, and the degree of willingness of the person expected to be put at higher risk. In some cases, such projects seem very straightforward: the mission is undertaken voluntarily, the level of risk is not too high, and the benefits are of great importance. In other cases, the risk may be low, the benefits great, but none is volunteering. Or, the risk may be very high but someone volunteers and the benefits are high. Yet, in other cases, the risk may be small and taken on voluntarily but the benefit is negligible. I suggest that none of these three aspects can by themselves serve as sufficient ground for justification for putting someone else at risk, but that they need to be addressed in some respect. It may be that in certain cases one of the aspects can be overlooked but that can only be if the other two are sufficiently outbalancing it.

## Compensation for Unequal Distributions

---

We now turn to the second question: how can an unequal but justified greater level of risk be duly compensated, or in other ways addressed, such that it can also be considered fair? Let us rephrase this question to a distinction between activities that are socially justified and those that are justified to those that are put at risk. This distinction was somewhat blurred in the previous section since part of what we take to justify an activity that poses greater risk on some, is to address the particular risks and benefits and consent of those at risk. Here we will look more closely at the high-risk group and the fact that they are put at greater risk.

As suggested in the previous section, the three reoccurring themes in the literature on risk is to look at consent, benefits, and level of risks. Some stress the level of risk, others the role of consent, and others still the role of benefits. I have suggested that we take all three into account as a triad when we assess risks. Now benefits and risks can both be adjusted to address the fairness of such greater risk for public benefits. We can add extra protective devices to counteract the level of risk directly targeted to high-risk groups. We can add extra benefits in terms of compensation to the benefits of particular risky activities. Both these aspects can thus increase the fairness of such otherwise unfair distributions of risk, and they may also affect the degree of its reasonableness and the extent to which it is agreed to by the people put to higher risk.

Keeping the triad of compensation, consent, and level of risk in mind, these can to some extent be used to replace the role of each other. Compensation can be used as a means to justify risk impositions when consent cannot be obtained (Nozick 1974). It can also be offered as

a means to persuade someone to willingly accept a risk or encourage them to take up a risk role. In such cases, compensation is a substitute for consent or forms part of the reasons for a person to consent. Compensation can also be offered as a means to acknowledge harm or risk *post ante*. In such cases, it serves as restoration for harm brought about as a result of risk. Compensation can thus stretch from merely being an acknowledgement of risk already having resulted in a harm to making a particular task with a certain risk into a good deal for the risk-taker when the risk is very small. The fact that there is compensation, or even that it is a generous one, must again be understood against the level of risk and the level of precaution. Imposing a risk with a generous compensation may still be an unfair way to impose high risks instead of implementing the appropriate safety devices if those are more expensive. A point to explore further is that of nonmaterial compensation. Historically there are cases where greater risk-taking has been compensated for not so much through monetary compensation but in higher social status or social recognition.

Freely accepted risks do not automatically make all kinds of compensation or safety precautions redundant. We would probably not consider as fair a differential treatment of workers in a dangerous industry where one person was freely volunteering to work without safety measures, while everyone else worked with such measures in place.

However, all that said, in contrast to the distribution of risk, the distribution of compensation and precautions seems to fall much more easily in the already well-developed theories of distributing goods. Here the extra point is made about how those aspects can be used to make unfairness in distribution a little fairer. Furthermore, the justification of a particular distribution of risk may depend upon the level of precaution and compensation.

## What Is a Fair Distribution of Risk?

---

What, then, is a fair distribution of risk? In this chapter, we have looked at several ways to approach this question. First of all we started with the presumption that fairness is a relative term. What is a fair share to me is determined by what others receive. In contrast to parallel treatments of fairness in distribution about goods, the problem lies not in the tension between scarcity and equal shares, but rather in the tension between risk reduction and equal shares. A too strict requirement on equality of shares could slow down the risk reducing aims since many reductions may come gradually and not be able to protect each equally. A further problem is that many reductions of risk would require that some individuals put themselves at greater risk for the safety of others. The problem of fairness of risk is then not so much one of equal distribution, at least not of risks, but one of balancing equal treatment and equal concerns with the overarching goal of protection against risk. Sources of risks can be distributed in more or less equal ways, dangerous missions can be equally or unequally taken on and more or less voluntary, and precautionary measures can target those at risk equally or unequally.

We investigated the idea that perhaps we could address fairness in terms of equal treatment for each person in terms of protection against risk. It was suggested that we need to distinguish between fairness at the level of distribution and fairness at the level of precaution. This distinction was further elaborated into a twofold approach to think about fairness and risk in general. On the first level, we seek justification for a particular spread and distribution of risk, including any inequalities it may impose. On the second level, we seek to as far as possible address the claims to fairness that the inequalities in the first distribution gives rise to in terms

of compensation and precautionary attention. It was suggested that greater risk imposed on someone needs both a general justification for that inequality and a specific justification in terms of consent, benefits, and level of risk to those at higher risk.

## Further Research

The whole field of fairness of risk is relatively underexplored. The current chapter has suggested that while it is hard to apply distributive reasoning onto chances of harm there are interesting parallels to explore when it comes to that which we can more easily distribute: the sources of harm and the resources to redress inequalities of risk in the form of precautionary measures. Much remains to be said about how we can combine the ideals of equality with the special requirements on risk reduction. Topics merely touched upon in this chapter that could be further developed include developing Rawls' difference principle, consent or voluntariness, and compensation for the particular problems of societal risk. Another topic worth exploring, merely mentioned in this chapter, would be to further Pettit's ideas about nonarbitrariness and non-dominion and exposure to risk.

## References

- Altham JEJ (1983–1984) Ethics of risk. In: Proceedings of the aristotelian society, New Series, vol 84, pp 15–29
- Ashford E (2003) The demandingness of Scanlon's contractualism. *Ethics* 113:273–302
- Ayer AJ (1972) Probability and evidence. Macmillan, London
- Broome J (1984) Uncertainty and fairness. *Econ J* 94:624–632
- Broome J (1990–1991) Fairness. In: Proceedings of the Aristotelian society, New Series, vol 91. Blackwell, pp 87–101
- Frankfurt H (1987) Equality as moral ideal. *Ethics* 98:21–43
- Hansson SO (2003) Ethical criteria of risk acceptance. *Erkenntnis* 59(3):291–309
- Hansson SO (2004) Philosophical perspectives on risk. *Techné* 8(1):10–33
- Hayenjelm M, Wolff J (forthcoming) The moral problem of risk impositions: a survey of the literature. *European Journal of Philosophy*
- Hooker B (2005) Fairness. *Ethical Theory Moral Pract* 8(4):329–352
- Lenman J (2008) Contractualism and risk imposition. *Polit Philos Econ* 7(1):99–122
- Mirrlees JA (1982) The economic uses of utilitarianism. In: Sen A, Williams B (eds) Utilitarianism and beyond. Cambridge University Press, Cambridge, pp 229–261
- Munoz-Dardé V (manuscript, forthcoming book) Conversations with a flattened amish farmer: risk and reasonable rejection. In: Muonz-Dardé (forthcoming) Bound together: how the political is personal
- Nozick R (1974) Anarchy, state and utopia. Basil Blackwell, Oxford
- Parfit D (2002) Equality and priority. *Ratio* 10(3):202–221
- Parfit D (2003) Justifiability of each person. *Ratio*, New Series 16:368–390
- Perry S (2007) Risk, harm, interests, and rights. In: Lewins T (ed) Risk: philosophical perspectives. Routledge, London/New York, pp 190–209
- Pettit Ph (1997) Republicanism. Oxford University Press, New York
- Railton P (1983) Lock, stock, and peril: natural property rights, pollution, and risk. In: Gibson M (ed) To breathe freely. Rowman and Allanheld, Totowa, NJ, pp 89–123
- Rawls J (2003) A theory of justice. Harvard University Press, Cambridge, MA
- Reibetanz S (1998) Contractualism and aggregation. *Ethics* 108(2):296–311
- Roeser S, Asveld L (2009) The ethics of technological risk: introduction and overview. In: Asveld L, Roeser S (eds) The ethics of technological risk. Earthscan, London/Sterling, VA, pp 3–10
- Scanlon TM (2000) What we owe to each other. Harvard University Press, Cambridge, MA
- Taurek JM (1977) Should the numbers count? *Philos Public Aff* 6:293–316
- Thomson JJ (1986) Rights, restitution, and risk. Harvard University Press, Cambridge, MA
- Timmermann J (2004) The individualist lottery: how people count, but not their numbers. *Analysis* 64(2):106–112



# 37 Intergenerational Risks

*Lauren Hartzell-Nichols*

University of Washington, Seattle, WA, USA

<b><i>Introduction</i></b> .....	<b>932</b>
<b><i>History</i></b> .....	<b>933</b>
Defining Intergenerational Risks .....	934
The Pure Intergenerational Problem .....	935
The Nonidentity Problem .....	937
Harm and Intergenerational Risks .....	940
Present Versus Future Harming .....	943
<b><i>Current Research</i></b> .....	<b>946</b>
The Limits of Cost-Benefit Analysis .....	946
Precautionary Principles .....	952
Intergenerational Justice .....	954
<b><i>Further Research</i></b> .....	<b>957</b>

**Abstract:** Intergenerational risks are intuitively defined as long-term threats of harm that will affect future people. The nonidentity problem, however, challenges our ability to accept this definition. This chapter offers an interpretation of intergenerational risks as threats of *de dicto* harm or as threats of harmful conditions (drawing on distinctions made by Casper Hare and Joel Feinberg, respectively). One of the challenges of intergenerational risks may be understood, following Stephen Gardiner, as the fact that it may never be a generation's interest to engage in an intergenerational cooperative scheme. This may make promoting intergenerational risks tempting when, for example, such risks involve a deferment of potentially harmful outcomes. A challenge for intergenerational ethics is how to understand which intergenerational risks are morally acceptable and which are not. Several (usually implicit) approaches to addressing intergenerational risks are discussed including economic approaches, precautionary principles, and intergenerational justice. The chapter calls for much more work to be done to better understand the challenges posed by intergenerational risks. Despite the fact that intergenerational risks abound in today's increasingly globalized world, we – as a society – do not yet grasp the complexity of intergenerational risks or how we should address them.

## Introduction

---

Most often when we think about risks, we do not think about especially long timeframes. We tend to think about activities and events that could have harmful effects during our lifetimes. But there are many activities and events that pose threats of harm over much longer timeframes. Greenhouse gas emissions, for example, pose long-term risks because of their delayed impact on the climatic system. It is long-term risks like these that are the subject of this chapter. Since long-term risks will primarily affect future generations the discussion focuses simply on intergenerational risks (hence the title of this chapter).

Intergenerational risks raise special theoretical and practical challenges because of the very fact that these risks span more than one generation. One potentially challenging feature of such risks, for example, is that those who might be affected by risky activities cannot even in principle participate in the regulation of the risky activities in question. Another challenging feature of intergenerational risks is that risky activities may in part determine *who* will be affected because such activities will usually affect the personal identity of future people. For these reasons, intergenerational risks present challenges not present in intragenerational risks. Yet these features are much more common than we may think. Many risks that at first appear to be intragenerational are in fact intergenerational. The prominence of intergenerational risks only highlights the importance of addressing questions such as how we should think about risky activities that might affect future generations.

In asking how risk is related to moral value, Sven Ove Hansson identifies three major approaches: “1. Clarifying the value dependence of risk assessments; 2. analyzing risks and risk decisions from an ethical point of view; 3. developing moral theory so that it can deal with issues of risk” (Hansson 2007: 21). This chapter will touch on all three of these approaches as they pertain to intergenerational risks, though in so doing it will identify the need for much more work in all three areas. This discussion will reinforce Hansson's ultimate conclusion that “Introducing problems of risk is an unusually efficient way to expose moral theory to some of the complexities of real life. There are good reasons to believe that the impact on moral philosophy can be thoroughgoing” (Hansson 2007, p. 33). The need for conceptual clarity

about the challenges posed by and ways of addressing intergenerational risks drives the philosophical perspective taken in this chapter.

Throughout the discussion the intergenerational risks posed by anthropogenic climate change are used as paradigm examples. After all, historic emissions of greenhouse gasses have put us in a position in which we are facing threats of harm and in which our current options are limited. Further, in continuing to emit such gases we impose threats of further harm on future generations. What climate policies we adopt will not only shape Earth's future climate but also who will live and under what conditions. During the industrial revolution people may not have known they were imposing threats of harm upon us – a future generation. From our perspective, however, their actions were risky because they not only emitted greenhouse gasses without an understanding of the consequences of such actions but also made choices that shaped the way people all over the world lived in ways that posed at least some clear (and other unclear) threats of harm. Climate change is therefore not only a paradigm example of intergenerational risk from a forward-looking perspective, but it raises ethical questions about, for example, the significance of the knowledge or lack thereof on the part of risk-takers from a backward-looking perspective.

In the following section, I consider interpretations of the challenges posed by intergenerational risks both from the perspective of the pure intergenerational problem (as introduced by Stephen Gardiner) and the nonidentity problem (most prominently attributed to Derek Parfit). This discussion first illustrates the challenges of intergenerational risks and then reveals that it may be hard for us to even conceptualize intergenerational risks as posing threats of harm to future generations, the definition that otherwise would seem most intuitive. In response to this challenge, I draw on the work of Casper Hare and Joel Feinberg and argue that intergenerational risks may be understood as posing threats of *de dicto* harms or as imposing threats of harmful conditions on future people.

After articulating the nature of the challenge of intergenerational risks, I discuss several ways of thinking about how we should address such risks. I begin by briefly considering economic approaches to addressing intergenerational risks, illustrating that cost-benefit analyses may leave out or obscure important normative considerations. I further argue that if cost-benefit analyses are to guide decision making about such risks, these should use a very low discount rate, if they should discount at all. I then go on to consider the ways in which precautionary principles, which push us to err on the side of caution, may be understood as guiding action in the face of intergenerational risks. Finally I discuss the relevance of work on intergenerational justice to intergenerational risks, highlighting specific examples from theorists working on climate change.

In the process of surveying these issues, this chapter will identify those areas in need of further research. Despite the fact that intergenerational risks abound in today's increasingly globalized world, we – as a society – do not yet grasp the complexity of intergenerational risks or how we should address them.

## History

The first thing to note about intergenerational risks is the fact that very little has been said about them, especially by philosophers. As Martin Kusch has said of risks more generally, "It is striking and perhaps surprising to note that very few contemporary strands of political philosophy

contain explicit prescriptions on how to deal with risk and uncertainty” (Kusch 2007, p. 131). And while there is a body of literature on intergenerational ethics and intergenerational justice, the relevance of these issues to intergenerational risk has been underexplored. A recent collection addressing philosophical perspectives on risk, for example, fails to explicitly address intergenerational risk at all (Lewens 2007). Given this, I will devote this section to exploring the nature of intergenerational risks so that I can discuss the relevance of work on intergenerational ethics and justice to intergenerational risks in the following section.

## Defining Intergenerational Risks

---

Before delving into a philosophical discussion of intergenerational risks, we must be clear how we define such risks. It is not obvious, however, what intergenerational risks are. Risks are generally understood as threats or possibilities of harm, though colloquially the term is sometimes applied to both negative and positive possibilities. Stephen Perry, who has worked on the ethics of risk, defines harm as “a relatively specific moral concept which requires that a person have suffered serious interference with one or more interests that are particularly important to human well-being, and which for that reason are appropriately designated as fundamental” (Perry 2007, p. 202). Perry argues that the imposition of risk, even of physical harm, is not a form of harm in itself (Perry 2007, p. 190). But it does not follow from this, he argues, that risky actions cannot be wrongful. Rather, this implies that we should understand risks as posing threats of harm. Intergenerational risks then pose threats of harm to future people, possibly in addition to existing people. Upon initial consideration then, intergenerational risks may be understood as threats of harm to future generations or which involve more than one generation. This chapter will focus on negative intergenerational risks, those that involve threats of harmful outcomes, but much of the discussion may be reinterpreted to apply to positive “risks” as well.

A more technical definition of risk comes from economics. Economists are very precise about defining risks, distinguishing *risk* from *uncertainty*, where the former is defined as randomness with knowable probabilities and the latter randomness with unknowable probabilities (Kight 2002). The distinction between risk and uncertainty may also be understood as being between measurable and unmeasurable uncertainties. On this description, intergenerational risks are those that will affect future generations and about which we know how probable the possible outcomes are. That is, intergenerational risks involve measurable uncertainties. But how often are we able to measure the likelihood of possible outcomes when these span more than one generation? I suspect the answer is not very often. While I will elaborate on why this is the case in what follows, intuitively I hope it is not hard to see why there are probably very few intergenerational risks on the economic reading of risk. There are simply too many unknowable or simply unknown variables and possible outcomes when it comes to (distant) future outcomes for us to usually be able to measure the likelihood of the outcomes of a risky activity or event.

On the one hand a chapter on intergenerational risk may therefore seem misplaced or misguided. It seems there will almost always be uncertainties that render intergenerational “risks” to be “uncertainties,” at least in the technical language most often used to describe risk. On the other hand, we need a way to talk about risks, such as those posed by climate change, that are in fact intergenerational. And when we think about intergenerational risks we are

most often thinking about uncertain threats of harm to future generations (though we may also think about risks taken by earlier generations that threatened temporally delayed harmful effects as well). As a starting point for a discussion of intergenerational risks, then, we need to expand the narrow economic definition of risks to include long-term, uncertain threats of harm. This means that as I use the term here, there is rarely such a thing as a “calculated intergenerational risk” because there will almost always be uncertainties given the temporal scale of intergenerational risks.

Climate change presents many paradigm examples of intergenerational risks, both from backward- and forward-looking perspectives. Climate change is spatially and temporally diffuse in both cause and effect (IPCC 2007). People all over the world have contributed and contribute to climate change through the emission of greenhouse gases (including through causing land use and other environmental changes). And people all over the world have been, are, and will be differentially affected by climatic changes in diverse ways. Because climate change effects are also diffuse in time, it is an intergenerational problem. Climate change effects are temporally delayed from the actions that caused them and they are long lasting, meaning past and present greenhouse gas emissions will affect many future generations of people. People from different nations and even different generations will also experience the effects of climate change differently. Many, but not all, climate effects will be harmful in a diversity of ways. Some individuals will lose their lives because of direct effects of climate change (such as from especially strong hurricanes). But climate change is having and will have not only direct effect on people through its physical effects; it is having and will have derivative social effects. Changes in agricultural productivity will affect both local and global economies. Health-care systems will be stressed in areas highly impacted by climatic effects. Lifestyle choices might become limited. Climate change clearly poses threats of harm to future generations, thus actions contributing to climate change contribute to the creation of these intergenerational risks. And past emissions of greenhouse gasses have affected and will affect relative future generations as well.

## The Pure Intergenerational Problem

Stephen Gardiner points out that there are severe moral problems that are conceived of in terms of generations, those posed by intergenerational risks being a prime example. He addresses the issue of temporal moral distance, identifying what he calls “the pure intergenerational problem” or PIP (Gardiner 2002, 2003, 2006a, 2011). Gardiner’s discussion of PIP helps make sense of the role generations play in ethics and as such is helpful to the present discussion of intergenerational risks and the challenges they present. Gardiner (2003) conceives of PIP by imagining a world with temporally distinct groups of inhabitants, each of which is solely concerned with its own, independent interests. He asks us to imagine that each group has access to goods that give their generation a modest benefit but which impose high costs on all later groups. PIP arises in this situation from the following dilemma:

- ▶ (PIP1) It is *collectively rational* for most generations to cooperate: (almost) every generation prefers the outcome produced by everyone restricting pollution over the outcome produced by everyone overpolluting.

- ▶ (PIP2) It is *individually rational* for all generations not to cooperate: when each generation has the power to decide whether or not it will overpollute, each generation (rationally) prefers to overpollute, whatever the others do (Gardiner 2003, p. 484).

This presents a rather intractable problem in that the asymmetrical position of the first group makes it rational to overpollute. This problem is iterated such that each subsequent generation will also be in a position where it is rational to overpollute. This problem is made worse by the fact that the description of this world rules out the possibility of reciprocity. It is never rational for a generation to be the first to cooperate. If there are ever moral reasons in favor of cooperation, which I suspect there are, PIP then presents a problem of fairness.

Gardiner (2003) identifies six features of PIP, which help us understand its relevance to talk of generations in ethics. The first feature is *temporal asymmetry*, which relates to the fact that groups are temporally distinct in the description of the problem. The actual world does not have this feature since generations overlap, though this concept helps define generations in the description of PIP. Second, there is *causal asymmetry* such that earlier groups can affect later groups, but not vice versa. This feature applies to the actual world as well and may imply that PIP has some practical application. Third, there is *asymmetric independence of interests* in PIP such that the interests of earlier groups are independent of those of later groups. While this is not strictly a feature of the actual world, it seems empirically likely that the interests of earlier people dominate those of later people in practice, making a degenerate form of PIP potentially applicable. Fourth, in the description of PIP groups are generationally *self-interested* (have a self-interest relative to their own generation), which, while not true of actual individuals, may partially describe the way people will act in time-indexed ways as generational groups in practice. Fifth, PIP involves *temporally diffuse goods* that offer modest benefits to a group while imposing costs on future generations, but other goods with deferred costs present similar challenges. Again degenerate forms of the problems presented by PIP arise even if the nature of the deferred costs is different. Finally, PIP has a *sequential* aspect that gives rise to its iteration aspect, and here again the details of this feature may be changed to similar effect.

Together these points help illustrate that many of the reasons for describing the groups in PIP as generations applies to talk of generations in the actual world, and many of the challenges presented by PIP apply in some degenerate form to intergenerational ethics. PIP most clearly applies to the actual world if we imply a wide definition of generations and include those future people whom the presently living will never meet. But Gardiner argues that many cases in which generations overlap will have much the same structure as PIP and thus present many of the same challenges. We can learn from PIP even if its claims are not all true in all cases.

One way of thinking about intergenerational risks is as the imposition of the possibility of harm on future generations by an earlier generation. The pollution example Gardiner uses to articulate PIP may be understood as an intergenerational risk since it involves one generation imposing the potentially harmful effects of their polluting activities on future generations. Even if all generations could agree that it would be best if no generation polluted, polluting activities in virtue of presenting intergenerational risks are attractive to whoever is polluting because the pollution's harmful effects are deferred. PIP helps us see why it may be tempting to impose such risks, especially when doing so incurs a benefit to the present generation. This is because intergenerational risks involve conflicts between the interests of generations since by definition such risks are taken by one generation to the potential detriment of others. Intergenerational risks therefore raise the questions of fairness that arise in PIP, though the

context in which such a risk is taken and the nature of the risk are certainly relevant to any detailed account of the fairness issues a particular risk raises.

In identifying climate change as presenting a perfect moral storm, Gardiner uses PIP to explain the intergenerational challenges or intergenerational storm of climate change (Gardiner 2006a, 2011). As we have seen, climate change will have potentially harmful effects that are significantly deferred. As such it presents intergenerational risks that can be understood as manifesting some form of PIP. Activities that contribute to climate change, such as the emission of greenhouse gasses and deforestation, are akin to the polluting activities in the description of PIP. From our perspective it may seem rational to engage in activities that contribute to climate change because of the very fact that the harmful effects of these activities impose risks on future generations rather than ourselves. But this quite clearly raises questions of fairness. It seems almost certainly unfair of us to impose risks on future generations when we know that our actions will have seriously harmful consequences. (Note that it is less clear whether it is unfair that earlier generations, who did not know what the consequences of their emitting activities would be, treated us unfairly when they emitted greenhouse gases.) Gardiner worries that the nature of PIP makes us vulnerable to moral corruption such that we may engage in manipulative or self-deceptive behavior as a means of avoiding accepting and/or addressing the complex moral obligations we face. This may mean that we will be able to act as if we are acting ethically when in fact we are avoiding addressing the full range of ethical implications of our actions and corresponding obligations. This is reinforced by our theoretical ineptitude with respect to how to address the ethical challenges of intergenerational risks, which comprises the theoretical storm of the perfect moral storm.

Understanding intergenerational risks through the lens of PIP helps us see that it is the very intergenerational nature of such risks that raises moral and practical challenges. Intergenerational risks by definition involve deferring potentially harmful outcomes on to future generations such that future generations may be unfairly disadvantaged by the activities of earlier generations. The fact that future generations cannot even in principle participate in the decision-making processes that lead to risks being imposed upon them may make it tempting for those considering engaging in risky activities to do so, even when such risks are unfair or unjust. A major challenge for intergenerational ethics is to identify when and why we ought not contribute to intergenerational risks, sometimes in the face of significant practical challenges.

## The Nonidentity Problem

Derek Parfit (1976, 1984, 2001) implicitly addresses the challenge of intergenerational risks by examining the different kinds of choices we have with respect to the future. Key to understanding the different kinds of choices we make is the assumption that the cells out of which a person develops determines who a person is. This view appeals to the *Time-Dependence Claim* that a person would have never existed if she had not been conceived when she was in fact conceived (Parfit 1984, p. 351). Two questions that Parfit asks to distinguish three different kinds of choices are: "Would all and only the same people ever live in both outcomes?" and "Would all and only the same number of people ever live in both outcomes?" (Parfit 2001, p. 295).

When the same people would live with either outcome, the decision is what Parfit calls a *same people choice*. In general, these are choices that do not significantly affect the future. A clear example of a same people choice is your decision to brush your teeth or not this morning. Whether you brushed your teeth or not mostly likely did not affect who will live in the future. When the same number but different people live with either outcome, the decision is a *same number choice*. An example of a same number choice is a woman's successful choice to use birth control to determine when she will have a set number of children. By using birth control a woman affects the personal identity of her offspring, at least in part, by determining when her children are conceived and hence their genetic makeup. Finally, when different people and a different number of people live with either outcome, the decision is a *different number choice*. These are choices that very much affect the future. A straightforward example of a different number choice is a woman's decision to have regular sex without using any form of birth control. Her choice affects how many children she will have, though she cannot predict the outcome of this choice in advance. (It is worthwhile to note that even the same people and same number choice examples given here could turn out to be different number choices if, for example, whether you brushed your teeth affected whether or not you had sex with your partner or if we consider the fact that children conceived at different times may themselves be more or less likely to have more or fewer children.)

Parfit argues that different kinds of choices require different moral considerations. He points out that we often apply our intuitions about same people choices to different people choices. This is not appropriate, however, because of the issue of personal identity. The fact that our personal identity is in part tied to our genetics has significant implications. Parfit poses a powerful question when he says: "It may help to think of this example: how many of us could truly claim, 'Even if railways had never been invented, I would still have been born?'" (Parfit 2001, p. 290). On reflection, the answer to his question for virtually all of us is "no." The invention of the railroad was significant enough to have affected the lives of the people who existed at the time and consequently who their offspring were, leading up to our current population. Had the railroad not been invented and introduced to society when it was, history would not have lead to our collective existence.

Climate change policies have the potential to be as or even more significant in the course of history as the invention of the railroad. Policies geared toward the development of new technologies could lead to advances or changes as or more significant than the development of the railroad. If this happens, future people 200 years from now will be able to answer the question "Even if such and such technology had never been invented, would you have still been born?" in the negative the way we today answer the question about the railroad. Similarly, if policies allowing the status quo to continue are adopted, future people 200 years from now will owe their existence to society's historical reliance on fossil fuels. Looking back in time we can also see that policies supporting the industrial revolution and the development of a fossil fuel based economy lead to our collective existence today. Not only would we not have existed if the railroad had not been invited, we most certainly would not exist had the industrial revolution not occurred.

By identifying the relationship between personal identity and choices that affect the future, Parfit exposes that fact that many decisions affect who will exist. When a decision affects who will exist, however, it is unlikely that different actions will result in the same number of people. Even if a decision causes the existence of just one extra person, or causes there to be one less person, it is a different number choice. Different number choices do not require a total change in the number and personal identity of future persons. When the quality of people's lives is

changed, the timing and parenthood of conceptions is changed. Therefore most of our choices that affect the lives of even small groups of individuals are different number choices. This means that most societal level choices are different number choices.

Casper Hare sums up this point nicely when he says, “Given that world history is a large and encompassing thing, it seems likely that most decisions that affect who exists will reverberate through it for many generations and unlikely that, when all is said and done, the numbers of people who ever exist will turn out the same whatever we decide” (Hare 2007, p. 520). Most, if not all, intergenerational risks involve different number choices. Acting in a way that imposes a threat of harm on future people is a different number choice in part because there is often uncertainty as to the outcome of such an action; this is part of what makes it an intergenerational risk. This implies that societal level choices that involve risk taking pose intergenerational risks, which in turn implies that very often when we are talking about risks we are talking about different number choices and hence intergenerational risks.

One of the most significant challenges posed by different number choices is the widely discussed nonidentity problem. The nonidentity problem is the problem that our choices can affect the future in normatively bad ways despite the fact that we cannot harm in causing to exist. Parfit states this as the problem of identifying the moral difference between outcomes that are worse for no one (Parfit 1984, p. 378). If we assume that causing to exist cannot be worse for someone who has a life worth living, then it is hard to see how choices that affect the personal identity of future people can be wrong, though they are worse for no one. The nonidentity problem is particularly challenging for different number choices, as most personal identity-affecting choices are, since in such cases it is hard to think about how to compare possible future populations of different sizes. It is hard to answer questions such as: Would it be better to have a large future population in which many people suffer? Or would it be better for there to be a smaller future population in which there is less suffering? In either case whoever lives owes their existence to our choices and hence as individuals cannot have been made worse off by our choices.

Parfit utilizes a generalized example of a policy decision involving a choice between depletion and conservation to illustrate and clarify the nonidentity problem (see Parfit 1984, section 123). I will use a more specific application of this example to the same effect that can help us understand its relevance to intergenerational risks. A nation must decide how it wants to regulate the use of fossil fuels. Initially, policy-makers take into account the fact that there are an infinite number of possible policies they could adopt that would result in an equally infinite set of possible outcomes. For the sake of example it is assumed that they have narrowed their decision down to the following three policy options:

- ▶ *Great Depletion:* This policy would allow unlimited use of fossil fuels until such materials become unavailable in 200 years. This policy will likely lead to a higher average quality of life in the short term, but a much lower average quality of life in the long term.
- ▶ *Lesser Depletion:* This policy would allow the use of fossil fuels, but would restrict this use more than the Great Depletion policy to extend the time fossil fuels would be available to 500 years. This policy will likely lead to a somewhat higher average quality of life in the short term, but a lower average quality of life in the long term.
- ▶ *Conservation:* This policy would greatly restrict the use of fossil fuels such that the use of fossil fuels would be distributed over a thousand years. This policy will likely lead to a lower average quality of life in the short term, but a much higher average quality of life in the long term.

We can assume that whichever policy we choose will affect population size, though we cannot know in advance in what way population size will be affected. The decision of which policy to adopt is therefore a different number choice. This is due to the fact that the different policy options will entail, among other things, different qualities of life for the citizens of the nation in which the policy is adopted, which we have already seen will affect the identity and number of future people.

Each of the different policy options will causally contribute to a different set of people living in the future. Although the future people in these possible futures would each experience different qualities of life, even when it is not assumed that causing to exist can benefit, it can at least be concluded that none of the policy alternatives are worse for anyone. The nonidentity problem arises when we nonetheless believe that there is a moral reason to adopt one policy over another. Parfit says, “The great lowering of the quality of life must provide *some* moral reason not to choose Depletion” (Parfit 1984, p. 363). He believes that the mere intuition that the Conservation policy should be morally preferred to the Greater or Lesser Depletion policies illustrates that the view that what is bad must be bad for someone must be rejected.

The challenge for intergenerational ethics is explaining why one policy is better than another when none of the policies would be worse for anyone. Key to the present discussion is that the nonidentity problem challenges the way in which we colloquially talk about harming future generations. If to harm someone is to somehow make her worse off or to negatively affect her interests, it seems we cannot harm future people whose existence we in part determine. The very definition of intergenerational risks as threats of harm to future people is threatened by the nonidentity problem since it is no longer clear that future people can be harmed at all.

One of the most significant aspects of the temporal delay between actions that contribute to climate change and their effects is the fact that these very same actions shape who will live in the future. Greenhouse gas emissions today will cause climatic effects in the future, but these actions will also simultaneously shape the future. What we do today with respect to greenhouse gas emissions policies and practices will affect what the future will be like. It will also affect who will live in the future. So in at least one important sense actions contributing to climate change are different number of choices insofar as these actions contribute to intergenerational risks. Similarly if we look back in time, greenhouse gas emissions helped shape what has become our present. Whether there was explicit acknowledgment of this fact or not (and at times there certainly was not while at others there was), previous actions that caused greenhouse gas emissions contributed to the intergenerational risks associated with the climatic changes we are experiencing today and which will continue into the future. Since we have already identified those actions that contribute to climate change as contributing to intergenerational risks, the question now is whether we can understand climate change as posing threats of harm to future generations, as our working definition of intergenerational risks suggests.

## Harm and Intergenerational Risks

---

Is there any means left for interpreting the risks posed by climate change or any intergenerational risks as being harmful? For identifying actions that further contribute to climate change as harmfully affecting future people? I think there is. But before offering an interpretation of intergenerational risks in response to the nonidentity problem, I should point

out that not everyone is so moved by this problem. Edward Page (2006), for example, after offering an in-depth discussion of the nonidentity problem and responses to it, in particular as these relate to intergenerational justice, takes the “no difference view” of the nonidentity problem. Much like Parfit, he argues that the “the problem should inspire us to think seriously about the theoretical basis for the responsibilities to which many of us are already intuitively committed” (Page 2006, p. 165). Nonetheless, in order to even make sense of intergenerational risks, given that these intuitively are threats of harm to future generations, we need a way to talk about harming relationships that span generations and the nonidentity problem challenges this possibility.

Casper Hare’s (2007) use of the *de re* (of the thing) vs. *de dicto* (of the word) distinction can help clarify how we may understand intergenerational risks without having to give up the notion that such risks involve threats of harm. Hare identifies two kinds of harm: one of which is the traditional kind of harm and one of which is essentially impersonal. He uses a helpful joke to illustrate this distinction: Zsa Zsa has found a way to keep her husband young and healthy. The joke is that the source of Zsa Zsa’s proverbial fountain of youth is that she gets remarried every 5 years. Zsa Zsa’s method for keeping her husband young and healthy promotes these good qualities in an impersonal *de dicto* way. Promoting these qualities in a *de re* way would require that she always have the same husband because *de re* goodness attaches to a particular person. While Hare’s paper focuses on *de dicto* goodness, it is *de dicto* badness or wrongness that is helpful for understanding intergenerational risks. Adapting Hare’s definitions we get:

- ▶ *De Re Worse*: Where S1 and S2 are states of affairs, S1 is *de re* worse for the health of \_\_\_ than S2, when the thing that is actually \_\_\_ is sicker in S1 than in S2.
- ▶ *De Dicto Worse*: Where S1 and S2 are states of affairs, S1 is *de dicto* worse for the health of \_\_\_ than S2, when the thing that is \_\_\_ in S1 is sicker in S1 than the thing that is \_\_\_ in S2 is in S2 adapted from Hare 2007, p. 514.

As Hare notes, and I agree, *de dicto* badness does not always matter, but the point is that it sometimes does. *De dicto* badness can apply to different number choices where “normal” *de re* badness cannot. As Hare says, “there are nonidentity cases . . . about which the *de dicto* betterness account gives clear answers, though they may involve actions that affect how many people ever exist” (Hare 2007, p. 521). Choices may be *de dicto* worse for future people even though they are different number choices.

A version of the classic Parfit example of two mothers who knowingly give birth to children with deformities or defects illustrates this point (see Parfit 1976). Imagine Katie conceives a child against the recommendation of her doctor while on a certain medication that causes her child to have a birth defect. Had Katie waited 1 month to conceive until she was off the medication, she would have had a normal child. Hare points out that we are right to think that Katie “makes things *de dicto* worse for the health of her future child, and this is something she should have been concerned to avoid” (Hare 2007, p. 516). Before deciding to conceive, there was no way for Katie to express *de re* concern for her child, since she could not at that point know the personal identity of her child. Yet it is appropriate to be concerned about one’s future children; this concern is a type of *de dicto* concern. Parents have especially strong *de re* obligations not to harm their children. But while parents cannot *de re* harm their possible future children, they can create or promote *de dicto* harmful conditions for

their future children. Katie's case is a clear example of this possibility. Ignoring her doctor's recommendation and trying to conceive while on the medication amounts to Katie's knowingly *de dicto* harming her future child, whoever he turns out to be. And this is a kind of *de dicto* wrong.

This distinction can help us understand the badness of climate change and other actions contributing to intergenerational risks in a way that avoids entanglement with the nonidentity problem. Actions contributing to climate change are (or contribute to) different number choices, affecting not only the personal identity of those who will exist but also how many people will exist. In terms of their interest-affecting climatic effects, such actions in contributing to the creation of future people cannot make these people *de re* better or worse. These actions can be, however, *de dicto* better or worse for whoever lives since these actions can impersonally affect the interests of whoever lives. It seems that in this case the way in which climatic effects can be *de dicto* worse for people matters morally speaking, though we still need an account of why this is so and what it says about the morality of climate-affecting actions.

While this helps us understand the badness of climate change we still need to explain the way in which intergenerational risks impose threats of harm on future people. My argument is that we cannot harm future people in the everyday *de re* sense, though we can do things that will be *harmful* to future people in a weaker, less direct sense. Namely, we can act in ways that will create *harmful conditions* for future people. We can *de dicto* harm future people. The climatic effects of actions contributing to climate change cannot *de re* harm, but these actions are harmful insofar as they can *de dicto* harm.

(Note that I am not claiming anything about responsibility here. It may turn out that nations, rather than individuals, should be held responsible for the *de dicto* harmful actions of their citizens. Sinnott-Armstrong (2005), for example, argues against individual responsibility. The present account of intergenerational risks does not immediately weigh in on the debate over who – nations, individuals, and/or other actors – should be understood as bearing responsibility for addressing climate change.)

Joel Feinberg's introduction to the concept of harmful conditions in the context of discussing harmless wrongdoing is helpful for further understanding *de dicto* harms. Feinberg says: "We can mean by the phrase *harmful condition* a state in which a person is handicapped or impaired, a condition that has adverse effects on his whole network of interests. By a *harmed condition*, on the other hand, we can mean a harmful condition that is the product of an act of harming" (Feinberg 1990, p. 26, original emphasis). An act can have harmful effects that do not actually harm by creating or promoting harmful conditions. The same kind of effect can be a harmed condition in one instance and a harmful condition in another. The key difference is the nature of the causal relationship between the act or actions in question, the effect(s) of this act(ions), and the person or people affected.

The harmful vs. harmed condition distinction helps clarify at least one key difference between contemporary and future effects of actions. An action puts one's contemporary into a harmed condition when it adversely affects her network of interests. But no present action can put a future person into a harmed condition since no present action can *de re* harm future people. Put another way, no future person will ever be able to rightly claim she exists in a harmed condition because of acts performed before her conception. Along these lines we can understand intergenerational risks as imposing threats of harmful conditions or *de dicto* harm on future people.

It is important to be clear about the context in which Feinberg uses the notion of harmful conditions, which is much more limited than the present context. Feinberg discusses the same Parfit example that Hare uses of a woman who wants to conceive but who is informed that if she conceives now she will have a child with a defective condition. If the woman waits 1 month for her temporary illness to reside she will conceive a child who will be completely healthy. In this case, when we consider the life of the “defective child” Feinberg points out that this child will not have been (*de re*) harmed should she come to exist. So long as her life is worth living, she will be able to say that she is glad her mother conceived her when she did, for if she had not *she* would not have existed, which would have been a much worse fate than living with her defect. Feinberg says of the state of the “defective child” if she is conceived, “I prefer to call it, therefore, a *harmful condition* rather than a harmed condition” (Feinberg 1990, p. 27, original emphasis). That is, he says she is not in a *harmed condition* because there was no prior act of harming that caused her to be in the state she is.

Feinberg says that cases of harmful conception are the only cases where a person is, “put in a harmful condition by the very act that brings him into existence, and the only example where determinations of harm require comparison of a given condition with no existence at all” (Feinberg 1990, p. 327). Feinberg appears to assume that the only cases where the same action that causes an individual to exist also causes her to live in harmful conditions will occur on the level of individuals conceiving or not conceiving in particular instances. While it is true that there is something special about cases in which the harm relationship is relative to the fact that the person would not exist were it not for the very thing that caused her to exist in harmful conditions, our actions also causally contribute to the existence of (distant) future people in ways that cause them to exist in harmful conditions where our actions are not those of harmful conceptions. When we contribute to an intergenerational risk we are potentially imposing harmful conditions on future people. Feinberg incorrectly concludes that the harmful conception cases are the only cases where this is so.

The implication of this is that Feinberg’s identification of the fact that we can create harmful conditions in which a person lives without and in spite of the fact that we cannot *de re* harm her is much more significant than he realized. It is not merely that so-called wrongful conception is harmful for the person it creates, our actions can be harmful to the people whose existence we causally contribute to, despite the fact that we cannot *de re* harm them. The important point is that Hare’s and Feinberg’s distinctions can help us see that there is a distinction to be made between *de re* and *de dicto* harming relationships, only the latter of which can hold between generations. *De dicto* harms impose threats of harmful conditions. This in turn can help us understand the nature of intergenerational risks as threats of harm to future people. Actions posing or contributing to intergenerational risks pose threats of *de dicto* harm to future people. And at least some of these threats of *de dicto* harm matter morally. The challenge is then to determine which intergenerational risks are morally acceptable and which are not. This is where we need to look to work in intergenerational ethics.

## Present Versus Future Harming

Before moving on to discuss how intergenerational risks may be addressed, it is important to recognize that there is a vast literature on the nonidentity problem. (One recent collection that

addresses this problem and points to the relevant literature is Roberts and Wasserman (2009). For non-consequentialist responses to the nonidentity problem see also Reiman (2007) and Kuman (2003). Here I will consider but one especially illustrative view that bears on the above distinctions so as to reinforce the interpretation of intergenerational risks presented here.

Elizabeth Harman (2003, 2004, 2009) has a very different view of what harm is than that which has been presented so far. On her view, we can in fact harm future people and not in a merely *de dicto* way. Harman asserts that a sufficient condition for harm is that, “An action harms a person if the action causes pain, early death, bodily damage, or deformity to her, even if she would not have existed if the action had not been performed” (Harman 2004, p. 107). On this account that something is a harm is a reason against it, but other reasons could outweigh the negative reason provided by the harm (Harman 2003, p. 116). Essential to Harman’s view is that benefits to future people offer reasons in favor of actions, but reasons against harm are “morally serious” and hard to outweigh (Harman 2004, p. 108).

The commonly used case of surgery illustrates Harman’s view. Suppose Susie is having her tonsils removed because she suffers from frequent bouts of tonsillitis. I, along with Feinberg and others, do not believe that the surgeon harms Susie when she cuts away her tonsils if the removal of Susie’s tonsils is in her interest. Harman, however, does believe the surgeon harms Susie when she cuts into her since in so doing she injures Susie’s body. Nonetheless, Harman would say that it is permissible for the surgeon to harm Susie in this way since she has good reasons for doing so, namely, to promote Susie’s overall health and well-being. So on Harman’s account this harm is not wrong. The implication of Harman’s view that is relevant to this discussion is that Harman argues that we can in fact harm future people and that the nonidentity problem is only a problem because it confuses what constitutes harming.

(I should note that in her 2009 paper, Harman discusses different number choices for which we do not know whether more or less people will exist depending on whether an action is performed or not. Most actions that pose or contribute to intergenerational risks are of this kind. Harman suggests that in such cases we will often have strong reasons not to act because, given the possibility that such choices will lead to fewer people existing, we should be cautious and refrain from acting. I do not find her discussion there particularly relevant to the present discussion, which is why I focus my discussion here on the implications of her general account of harm.)

Critical to Harman’s argument is her position “that an action may be wrong *in virtue of harming* even though it makes a person better off than she would otherwise be” (Harman 2003, p. 105, original emphasis). She articulates this more clearly when she claims that reasons against harm are so morally serious that the mere presence of greater benefits to those harmed is not in itself sufficient to render the harms permissible: When there is an alternative in which parallel benefits can be provided without parallel harms, the harming action is wrong (Harman 2004, p. 93).

Looking at some of the cases Harman uses to support her position helps to clarify where I believe her account goes wrong. One case Harman considers is that of a woman whose life is better because she had a child conceived as the result of a rape in part because she is remarkably able to separate the trauma of the rape from her attitude toward her child such that they have a normal parent–child relationship (Harman 2003, p. 103). A second case Harman discusses originally comes from James Woodward (Woodward 1986, p. 809); Viktor Frankl believes his imprisonment in a Nazi concentration camp actually made his life better than it would have been had he not been imprisoned because it enriched his character and deepened his

understanding of life. Harman further supposes that Frankl does not wish that the Nazi's had not imprisoned him, despite the fact that they clearly impermissibly harmed him (Harman 2003, p. 104). Harman rightly claims that both the raped woman and Frankl are better off because of the very acts that harmed them insofar as these events were pivotal life-experiences that lead them to be better off than they were before they were harmed. Harman is also correct to say that both the raped woman and Frankl need not regret that they were harmed for these harms to have been impermissible and for the victims to have legitimate complaints against those who harmed them. But, I think that Harman's discussion of these examples and their implications, while valid for the cases she discusses, is misleading.

The very nature of rape and imprisonment in a Nazi concentration camp clearly identify these actions as harms on any reasonable account of harm. I do not think anyone would disagree that at the time, the act of raping harmed the woman in question. And no one would disagree that Frankl was harmed by his imprisonment. Being raped or unjustly imprisoned are both physically injurious, damaging to one's interests, and a violation of one's rights. But not all acts of harm prove to be bad for us in the long run. We do not necessarily regret all the harms we have suffered despite having legitimate claims against those who harmed us. This is why perspective matters.

Acts of harm can be direct and immediate between persons: the rapist harmed the woman through the act of raping her. The harm incurred by an act can also be temporally delayed from the moment of harming: a rape victim could suffer psychologically days, weeks, or even years after the immediate harm such that they experience harm long after the physical act and a car accident victim could have a physical injury that does not manifest itself until hours after being hit by a drunk driver. Whether we regret being harmed in a given instance and whether it ultimately makes us better off depends on how we deal with being harmed and on the many other things that happen to us in our lives. Harman's rape victim was clearly impermissibly harmed since the act of rape not only violated her rights and injured her body but also thwarted her interests at the time. The fact that she later was glad for having been raped because of both its consequences and her choices says something about who she is and later occurrences in her life; it does not say anything about the impermissibility of the rape itself. Rape is always harmful; it is never in one's interests to be raped.

The significant point I want to make is that Harman's position about something being wrong in virtue of harming despite making a person better off than she would otherwise be does not apply to future people. Just because a person who has been harmed by a contemporary can later not regret this impermissible act of harming, and in fact may have benefited from it in the long run, does not mean that future people who would not regret the alleged harms against them would be in fact harmed. A future person whose interests are negatively affected by and/or is injured by the effects of climate change will almost certainly be glad that she came to exist, but she cannot believe that actions that causally contributed to her existence made her worse off because the alternative would have been her nonexistence, not her being in some better state. But this does not mean that she has been harmed by those actions in the same way the rape victim and Frankl were harmed. The rape victim and Frankl already had interests and bodies at the time they were harmed. So, while later they come to believe that the acts that harmed them were good for them despite having been impermissible harms, the reason they understand these acts to have been impermissible is in part because they were not necessary to their existence. The future person, however, cannot make such a claim. The acts that led to her interests being negatively affected and/or her being physically injured *had* to have happened in order for her to even exist and have a body and interests in the first place.

Harman tries to avoid the nonidentity problem by making the point of comparison for determining when something constitutes a harm as an ideal healthy bodily state. In so doing she avoids the problem that future people cannot be harmed because they cannot be made better or worse off, as is required to judge whether something is a (*de re*) harm or benefit. For Harman, *any* act causing bodily injury is an act of harming. It follows that any act that causes a future person to suffer a bodily injury would thus also be an act of harming. But, any act that causally contributes to a future person's existence also causally contributes to the creation of his body. It is hard to understand how from the perspective of any future individual, even if such an act also causes his body to be injured, it could be deemed wrong by Harman since it seems the goodness of the creation of the future person's body will always outweigh any later injury to it. So long as a person's life is worth living, as I assume virtually all people's lives are, the creation of his body will outweigh any wrongness of injuries to it caused by the same acts that created it. While Harman can say that a future person has been harmed on her view when he suffers bodily injury, none of these harms will be able to be judged as wrongs when judged independently. To me this takes away a key part of the meaning of "harm." This is why it makes more sense to focus on the way in which such actions are harmful such that we need distinct ways of thinking about intergenerational ethics. The clearest way to make sense of the way in which intergenerational risks pose threats of harm is by understanding the way in which such risks may create harmful conditions or constitute *de dicto* harms despite affecting the identity of future people.

## **Current Research**

---

Having now established a way of making sense of intergenerational risks, in this section I will discuss several approaches to addressing intergenerational risks. My intention is to survey work relevant to understanding how intergenerational risks should be addressed. In order to address intergenerational risks we must decide what (intergenerationally) risky activities are acceptable. But since very little has been written on intergenerational risks directly, I will focus on aspects of the literature that I find to be particularly relevant and/or promising for being able to contribute to a better understanding of the relevant issues. In what follows I will first consider the way in which economics addresses intergenerational risks and the use of discounting in cost-benefit analyses in particular. I will then discuss the precautionary principle as a possible guide to addressing intergenerational risks. Finally I will briefly look to the literature on intergenerational justice, using examples from work on climate change, to illustrate the potential applicability of work in this field to the ethics of intergenerational risks. This discussion will illuminate the fact that thinking about how to address intergenerational risks pushes us to think deeply about intergenerational ethics. What we think we owe future people, after all, affects what risks we believe we can permissibly impose upon future generations.

### **The Limits of Cost-Benefit Analysis**

---

It is especially important that we consider what is probably the dominant approach to addressing intergenerational risks, namely, economic approaches. When faced with a threat of harm we often perform cost-benefit analyses to determine an appropriate course of action.

There are two primary ways, however, in which cost-benefit analysis is understood. First, cost-benefit analysis is sometimes understood as a method for analyzing quantitative economic information. Understood as a quantitative method, cost-benefit analysis is merely an analytic economic tool, a quantitative method that produces quantitative results. Second, cost-benefit analysis is most often understood as a decision-making framework that makes policy recommendations based on assessments of the costs and benefits of different policy options. This distinction is supported in the literature, though it is not consistently discussed or identified in exactly the way I describe. Many authors have noted that what is implied by “cost-benefit analysis” varies widely from context to context. As Richard Posner says, at the highest level of generality this term is used synonymously with welfare economics (2001). At the other end of the spectrum it is meant as a principle of wealth maximization. Both of these interpretations identify cost-benefit analysis as a decision-making framework, but they associate cost-benefit analysis with different decision rules. Posner, however, also identifies the interpretation of quantitative cost-benefit analysis as “a method of pure evaluation, conducted wholly without regard to the possible use of its results in a decision” (Posner 2001, p. 318).

When understood as a framework for rational decision making, cost-benefit analysis is associated with certain norms. The most general form of the normative cost-benefit analysis decision rule, or cost-benefit principle as it is sometimes called, tells us to take the action with the greatest expected net benefit. The rule is also sometimes stated such that one ought to choose the project that maximizes the present value of net benefits, where the net benefits of projects are generally equated with the aggregate welfare consequences of those involved in or affected by the activity under consideration. This rule is also often understood as saying that benefits should be maximized and costs minimized, or that benefits should be maximized in the most cost-effective way possible. The welfare consequences in cost-benefit analyses are usually measured by individuals’ willingness-to-pay.

The first thing to note about cost-benefit analysis is that it may leave out and obscure important normative considerations. First, cost-benefit analysis may be misleading insofar as what gets counted as costs and benefits involves making both normative and pragmatic judgments. Cost-benefit analysis requires assigning a monetary value to everything it evaluates. But what costs and benefits get accounted for in cost-benefit calculations is often in part determined by what is readily monetized, which can lead to both errors (i.e., assigning wrong values) and omissions (i.e., excluding things of value). Second, cost-benefit analysis can appear to be an objective or scientific method for making decisions when it is in fact not. The clear procedural structure of cost-benefit analysis obscures both the fact that a range of normative judgments underlie the quantitative information that it assesses and that it sometimes fails to consider important normative issues.

Imagine a cost-benefit analysis of the decision to log an area in the Amazon. What is the monetary value of a hectare of the Amazon rainforest? Economists often quantify this in terms of how much money could be gained by logging the area, but clearly there are other ways in which a hectare of rainforest is valuable, some of which are more difficult to quantify than others. Instrumentally a hectare of rainforest is valuable to humans in all kinds of ways. Locally, the rainforest provides food, firewood, and other resources to its residents. Globally, the Amazon plays a role in the climatic system and would contribute to climate change if it were deforested. But how much monetary value should we assign to the costs associated with deforesting a hectare of the Amazon? Tropical rainforests also contain plants with potential medicinal value. If we cause these species to go extinct, we lose the potential to benefit from

their medicinal uses. But how much value does this add to the rainforest quantitatively? There are also people who believe that the rainforest is valuable for its own sake, that it is intrinsically valuable. How could the intrinsic value of the rainforest and of the species that exist there be included in cost-benefit calculations? And how do we incorporate or assess the long-term instrumental and/or intrinsic value of the Amazon?

It is difficult to imagine that any cost-benefit analysis involving the Amazon could fully capture the full value of this resource. What costs and benefits are quantified will depend both on pragmatic limitations and normative judgments about what aspects of the Amazon are valuable and how these values should be monetized. The results of such cost-benefit analyses will therefore be influenced by what were considered and quantified as costs and benefits and will obscure possible impacts and outcomes that never made it into the calculations. This may be particularly important if we are trying to assess and make decisions about intergenerational risks that may have long-term effects.

The first point about the value-ladenness of cost-benefit analysis is therefore that how costs and benefits are measured involves value judgments because of the subjectivity of the metrics used. The willingness-to-pay metric usually used to measure costs and benefits depends on the values and preferences of individuals. How much individuals are willing to pay to protect the environment, for example, will likely be informed and influenced by their preexisting normative stances such as how much they value the environment as a means (e.g., for recreation) or for its own sake. One person might judge the paving of a wetland for the development of a shopping center as a benefit if they dislike wetlands for aesthetic reasons, whereas someone else might judge this as a cost because they value natural areas and the species that live in wetlands. Uncertainty further complicates and distorts willingness-to-pay assessments. Often people do not have enough information to fully understand and assess all possible outcomes even when such information is available.

Any objectivity in cost-benefit analysis comes from the quantitative methodology it uses. But it is important to recognize, as most people do, that the inputs into cost-benefit calculations can be subjective and/or normative. (I should note that the use of willingness-to-pay as a key measure of value in cost-benefit analysis is the basis of many objections against cost-benefit analysis. I too worry about the use of this metric, but I will not pursue this issue at great length here.) The very fact that quantitative cost-benefit analysis is quantitative can make it appear to be a value-neutral methodology for assessing costs and benefits. But since what gets counted as costs and benefits in the first place requires normative judgment, even quantitative cost-benefit analysis is not value-free. The quantitative results of cost-benefit analyses must therefore be contextualized to the assumptions about what was measured and assessed. Further, the decision rule that guides cost-benefit analysis can appear to be more objective than it is because of its clear procedural structure (e.g., choose the option with the highest benefits and least costs). But underlying this decision rule is the valuing of efficiency (or whatever is being maximized or captured by the decision rule).

Finally, when cost-benefit analyses use willingness-to-pay as a metric in their calculations they may implicitly ignore considerations of equality and justice. This is especially true of wealth-maximizing versions of cost-benefit analysis because these leave out considerations of distributive justice and, as Posner says, “it treats a dollar as worth the same to everyone” (Posner 2001, p. 318). The problem is that gaining or losing a dollar has a much greater impact on a poor person’s welfare than on a wealthy person’s. If the preferences or utility of rich people or states are weighted more heavily in a cost-benefit analysis, the rich are likely to also

benefit the most from any policy the analysis recommends. For example, if the policies being evaluated are about national park access fees and it is determined that the rich are willing to pay more than the poor, then the results of a cost-benefit analysis might show that it would maximize net benefits to charge entry fees that are prohibitively expensive for the poor. Such a policy, however, may violate principles of justice by essentially denying the poor access to a public good. Obviously there would be many factors to consider in a full cost-benefit analysis, but there is the potential here for distributive injustices to be exacerbated. Using willingness-to-pay may not only treat rich and poor people unequally but may also result in recommendations that exacerbate injustices between the rich and poor. Furthermore, when cost-benefit analyses are done on a global level, they can, as Amartya Sen says, “obscure an enormous issue of justice and fairness between different parts of the world” (Sen 1995, p. 32). More importantly for this discussion, using willingness-to-pay as a metric may fail to appropriately capture long-term value, which it seems we should be concerned about when considering intergenerational risks.

The second thing to note about cost-benefit analysis has to do with its methodology. Discounting is used in economics, and specifically in cost-benefit analysis, to account for changes in the value of different metrics over time. Using a positive discount rate in economic analyses assumes that wealth will continue to increase over time and that relative values decrease over time. There are two reasons we may want to privilege a low discount rate when assessing intergenerational risks. First, this enables all possible outcomes to be considered whenever they threaten to occur. Fully assessing intergenerational risks requires us to be temporally neutral in this respect, which means we should grant future people equal ethical standing with respect to the harmful outcomes they might experience. Second, discounting the future has the potential to mask potentially harmful outcomes. A complete assessment of intergenerational risks therefore may require assessing all possible future impacts and using a zero or near-zero discount rate. Discounting the future has the potential to mask the harmful effects of intergenerational risks that should, after all, be the focus of an analysis of such risks. These points could be understood as applying to the two aspects of the discount rate, the rate of pure time preference and the rate of consumption, respectively. Treating all generations as having equal ethical standing requires using a zero (or near zero) rate of pure time preference, while taking into account the possibility of thresholds and non-compensate-able harmful outcomes requires using a low rate of consumption (which otherwise assumes future people will be wealthier and therefore benefit less per unit of consumption). Considering what this means for an economic analysis of the intergenerational risks climate change presents will help make this point clear.

The *Stern Review* of the economics of climate change argues that the approach to discounting taken in an economic assessment of climate change must diverge from traditional approaches because such an approach “must meet the challenge of assessing and comparing paths that have very different trajectories and involve very long-term and large inter-generational impacts” (Stern 2007, p. 25). Different climate change policies (different emissions scenarios) could lead to drastically different outcomes. The difference between high-emission scenarios and the lowest possible emission scenarios is a difference of many degrees in terms of global average temperature. Because of this, the *Stern Review* takes into account ethical arguments in determining an appropriate discount rate. The most influential ethical assumption the *Stern Review* makes is that, “if a future generation will be present, we suppose that it has the same claim on our ethical attention as the current one” (Stern 2007, p. 35). The *Stern Review* argues that this implies that treating the welfare of future generations as on par with our own because the prospects of

future generations matter follow from “most standard ethical frameworks” (Stern 2007, p. 37). This follows from John Broome’s conclusion in his contribution paper that he sees “no convincing grounds for discounting future lives,” which seems appropriate in the context of the assessment of intergenerational risks (Broome 2006, p. 19). (See Caney 2008, 2009 for a related argument that a human rights-based account requires that we do not have any pure time preference.) This stance about the ethical standing of future people leads the *Stern Review* to use a very low rate of pure time preference of 0.1% (Stern 2007, p. 663). (The reason the *Stern Review* has a positive rate of pure time preference at all is to account for the possibility of human extinction. There is a genuine question as to whether the *Stern Review* is right to account for the possibility of human extinction, but exploring this issue is outside the scope of this chapter. The *Stern Review* therefore generally uses a discount rate of 1.1%, though the *Stern Review* says that it does not always use a consistent discount factor and thus discount rate, see Stern 2007, p. 60. See also Dietz et al. 2007 for a reiteration and further defense of the *Stern Review*’s treatment of discounting.)

William Nordhaus, on the other hand, consistently uses a relatively large discount rate in his analysis while admitting that the choice of the discount rate is especially important for climate change because it will significantly impact the future – including the far future. He says, “The approach in the DICE [Dynamic Integrated model of Climate and the Economy] model is to use the estimated market return on capital as the discount rate. The estimated discount rate in the model averages 4% per year over the next century” (Nordhaus 2008, p. 10). (See also Nordhaus 2008, Chap. 9.) Nordhaus illustrates that according to this rate this means that \$1,000 of climate damages 100 years from now is valued at \$20 today. It is therefore unsurprising that Nordhaus has criticized the *Stern Review*’s “extreme” assumptions economic discounting, which results in its “radical view of policy” (Nordhaus 2007a, p. 689). He says, “The *Review* seems to have become lost in the discounting trees and failed to see the capital market forest by overlooking the constraints on the two normative parameters” (Nordhaus 2007a, p. 700). Despite the fact that Nordhaus believes that economic analysis should not be the sole guide to decision making, he clearly has strong opinions about how economic analysis should be done – using a relatively high discount rate.

Nordhaus argues that the *Stern Review*’s results stem from its use of a low time-discount rate and low inequality aversion (Nordhaus 2007b, p. 202). He believes the near-zero pure rate of time preference and correspondingly low discount rate used in the *Stern Review* magnifies “large and speculative damages in the far-distant future” into a large current value (Nordhaus 2007a, p. 696). Nordhaus further demonstrates, through a series of basic calculations, that if the *Stern Review*’s parameterizations are corrected to match standard economic methods and assumptions, in part by using a higher discount rate, the results are in line with standard economic models (such as his own) (see Nordhaus 2007a: Section 4, 697–701). He calculates that more than half of the estimated damages “now and forever” cited by the *Stern Review* occur after the year 2800, which he implies should be taken as reason against the review’s use of an extremely low discount rate (Nordhaus 2007a, p. 696). Martin Weitzman also estimates that the difference between the *Stern Review*’s discount rate and more traditional discount rates changes the estimated damage costs of climate change 200 years from now by two orders of magnitude (Weitzman 2007, p. 708). The question we must ask is whether it is in fact problematic, as Nordhaus suggests, that the *Stern Review* incorporates climate change effects in the (relatively) far-distant future. In the context of thinking about intergenerational risks I think it is not.

There are at least two main problems with Nordhaus' arguments in favor of using a "standard" discount rate and making recommendations based on these calculations. First, Nordhaus does not sufficiently take into account the ethical standing of future people. When it comes to addressing uncertain threats of harm that have the potential to affect future generations, the *Stern Review* makes ethical assumptions that enable it to more fully assess and therefore make more appropriate recommendations about the intergenerational risks climate change poses. But while the *Stern Review* focuses its arguments in favor of using a low discount rate on this point, there is a second reason why Nordhaus' arguments are misguided and which maybe even the *Stern Review* does not sufficiently consider. This second problem is that climate change does not fit the assumptions made by standard economic approaches insofar as there might be thresholds after which we will not be able to avoid potentially harmful climate impacts and climate change threatens to have harmful effects that future people may not be able to compensate for. Eric Neumeyer (2007) makes a similar argument that even the *Stern Review* fails to adequately address the irreversible and nonsubstitutable damages and loss of natural capitol that climate change will incur, though he makes this argument in contrast to arguments about discounting. He argues that incorporating the importance of such damages would have provided an even more compelling case for drastic action to mitigate climate change. Traditional cost-benefit analysis fails to account for the possibility that preventing some threats of harm cannot be delayed and that some harmful outcomes cannot be compensated for once they occur or are initiated.

Atmospheric greenhouse gas concentrations take a long time to dissipate, which means that there will be a delay between our acting to reduce greenhouse gas emissions and the climatic impacts (e.g., temperature changes) that will occur. The longer we delay action to reduce greenhouse gas emissions, the higher global average temperatures are predicted to rise. The problem is that global average temperatures will not immediately decrease if we later decide we want to take action to reduce atmospheric greenhouse gas concentrations. For example, once the West Antarctic Ice Sheet melts, we will not be able to refreeze it or prevent sea level from rising, no matter how wealthy we are. There are therefore thresholds or points-of-no return for preventing certain climate impacts. Using a high (or even any positive) rate of consumption in the discount rate does not sufficiently account for this because it assumes that future people will be able to use their increased wealth to address any harmful climate impacts. But, future people will likely not be able to compensate for the five or more meters of sea level rise that is predicted to occur if the West Antarctic Ice Sheet melts (IPCC 2007, p. 819), no matter how wealthy they are because the extent and magnitude of the impacts would simply be too great. We will not always be able to compensate for harmful outcomes, since, for example, some harmful outcomes could undermine our economic system or be so costly as to be beyond compensation.

These and other reasons may lead us to believe that we should use a very low discount rate, if we should discount at all. However, as Gardiner points out, rejecting discounting need not mean that we use a uniform zero discount rate (2011). We need to understand what the implications of our policy options are both now and in the future so that we can decide what the most appropriate course of action is. While I do not deny that future people will sometimes be able to compensate for some climate impacts, for example, by developing technologies to prevent such impacts from being harmful, we cannot *assume* that this will always be the case when we are talking about widespread, irreversible changes. In trying to understand how to address intergenerational risks we must understand in what ways future people will be affected

and whether or not they will be able to compensate for potentially harmful effects. We will not understand the implications of intergenerational risks unless we draw on analyses that significantly discount the future when assessing such risks. If economics is to guide decision making about intergenerational risks, cost-benefit analyses assessing such risks should use a very low discount rate, if they discount at all. Careful consideration of the normative assumptions guiding the quantification of the inputs to such cost-benefit analyses is also needed. Intergenerational risks raise ethical challenges that standard economic approaches may fail to account for.

## Precautionary Principles

---

An entirely different approach to addressing intergenerational risks would be to apply the precautionary principle. The precautionary principle has been used in many contexts and has many formulations. The general idea behind this principle is that it is better to be safe than sorry and/or that we sometimes ought to act in advance of scientific certainty. One way of understanding the precautionary principle is as guiding us toward caution in the face of risks, including intergenerational risks. As such it is at least implicitly one of the most discussed principles for addressing intergenerational risks. Nevertheless, there is no consensus as to how the precautionary principle should be formulated or interpreted. And if the precautionary principle is to guide decision making about intergenerational risks, we must be clear as to what this principle requires.

The first set of challenges the precautionary principle faces is that its components are often not clearly defined or delineated. These challenges concern the content of the precautionary principle. Derek Turner and I argue that the precautionary principle is inherently unclear in five ways: first, the precautionary principle is often formulated such that it is unclear who must take responsibility for and bear the cost of precaution; second, the precautionary principle is unclear and even internally contradictory when it applies simultaneously to threats of harm to human health and the environment; third, it is often unclear what are to be counted as threats of harm; fourth, it is also unclear what are to be counted as precautionary measures (see also Sandin 2004, 2007); fifth, it is often unclear how much precaution is required (Turner and Hartzell 2004). Of particular importance is a variant of the third issue: risks abound on all sides yet we cannot possibly be precautionary about everything (Sunstein 2005 and Posner 2004 also state versions of this objection). I identify this as the paralysis objection; the precautionary principle would be paralyzing if it were to require precautionary measures against any and all threats of harm.

Another set of challenges facing the precautionary principle stems from confusion about just what this principle is supposed to be. Not only does the precautionary principle appear to have many formulations, but also its very nature seems to vary depending on the author and situation. First, the precautionary principle is sometimes understood as a family of principles that share a common structure (see, e.g., Sandin 2007; Manson 2002; Hickey and Walter 1995). Individual formulations of the precautionary principle, according to this view, may be tailored to specific applications. Several authors have suggested features that such specific formulations of the precautionary principle must share (see, e.g., Manson 2002; Hickey and Walter 1995; see also O'Riordan and Jordan 1995). Other authors who ascribe to this view have emphasized

that there are some particularly plausible versions of the precautionary principle (see, e.g., Gardiner 2006b). Second, the precautionary principle has been interpreted as an approach to risk management such that it is more of a decision-making procedure or framework than a principle (see, e.g., Goklany 2001; Resnik 2003). On this view, the precautionary principle is seen as a guide to rational decision making in the face of uncertain threats of harm. A third view of the precautionary principle is captured by Kerry Whiteside's conclusion that "the precautionary principle shades off into precautionary politics" (Whiteside 2006, p. 150). On this view, the precautionary principle is understood as a banner for a new way of thinking about risk management in the face of uncertainty. Along these lines Turner and I note that the precautionary principle has come to serve as a banner signifying a shared commitment to the welfare of the environment and future persons, and in addition shared reservations about the effectiveness and applicability of economic cost-benefit analysis (Turner and Hartzell 2004; O'Riordan and Jordan 1995 make a similar point). Whiteside understands the precautionary principle as guiding us where standard assumptions about risk management do not hold by shifting the default position toward action rather than business as usual in the face of uncertain threats of harm. But he argues that there can be no 20- or 30-word juridical formulation of the precautionary principle because what is meant by "the precautionary principle" is this shift in the way we approach decision making.

Elsewhere I have brought attention to the apparent dual roles of the precautionary principle; it appears to call for precaution in the face of uncertainty as well as a new way of thinking about decision making (Hartzell 2009). I argue that rather than thinking of the precautionary principle as a loose family of principles or as a decision-making procedure, we should require all precautionary principles to be understood as individual moral principles capturing particular *prima facie* moral duties. This means there may be other moral duties that will sometimes supersede precautionary principles, but we have strong moral reasons to follow precautionary principles. We should require that all precautionary principles bear some connection to the intuition that it is better to be safe than sorry and/or to the idea that we should sometimes act in advance of scientific certainty, but this alone is not enough. If a precautionary principle is to have any real normative force it must pick out a unique *prima facie* moral obligation. All things being equal, in some (but certainly not all) cases we *ought* to take a better safe than sorry approach and act in advance of scientific certainty. There is thus a family of precautionary principles, but each precautionary principle is an independent moral principle that must be independently identified and justified. This understanding of precautionary principles clarifies that "the" precautionary principle cannot be used to justify any and all precautionary action. It returns meaning and normative force to precaution in the limited cases in which precaution is morally called for.

Precautionary principles then may guide decision making about intergenerational risks, but it is misguided to think that "the precautionary principle" can be helpful as there is no such single principle. One of the looming questions about intergenerational risks, after all, is how to determine when they are morally acceptable and when they are morally impermissible. A simple notion of "better safe than sorry" or "act in advance of scientific certainty" would push us to be extraordinarily averse to intergenerational risks since there will almost always be uncertainty about the possible outcomes of our actions. But limited precautionary principles may be able to guide us in the face of such uncertainty and help us sort out when we may and when we ought not to impose threats of harm upon future generations.

## Intergenerational Justice

---

The third established body of work that may help us understand how to address intergenerational risks comes from the study of intergenerational justice. Work on intergenerational justice applies to intergenerational risks insofar as such risks pose questions of justice. One question we may ask is, when it is permissible and impermissible to impose risks on future generations? Which intergenerational risks are morally acceptable and which are unjust? Few authors who have written on the topic of intergenerational justice have explicitly addressed intergenerational risks. But I see work on intergenerational justice as providing an insightful way of understanding the ethics of intergenerational risks. I do not have the space here to provide a thorough treatment of all work on intergenerational justice, but I hope that by offering a few examples from work on climate change this section will illustrate what the field of intergenerational justice has to offer to an understanding of intergenerational risk and why the significance and prevalence of the latter in many ways begs for more work on intergenerational risk. (A recent anthology, *Intergenerational Justice*, illustrates the breadth of work in this field that goes far beyond work on climate justice (Gosseries and Meyer 2009).)

If we understand climate change as involving intergenerational risks such that actions contributing to climate change pose time-delayed threats of (*de dicto*) harm, then we can see work on climate justice as trying to make sense of the ethics of intergenerational risks. And if we assume, as most justifiably do, that at least some of the intergenerational risks posed by climate change are morally unacceptable, thinking about what justice requires in terms of eliminating or reducing these risks, via mitigation and adaptation, can help us move from working to understand the nature of such risks to understanding how such risks should be addressed. Edward Page says, “The use of the language of entitlement and justice is important: although there is a range of alternative reasons for preserving environmental goods for present and future generations, justice appears to provide a more compelling and urgent defense of environmental and intergenerational duties than rival approaches” (Page 2006, p. 162). For all of these reasons intergenerational risks lend themselves to an intergenerational justice approach.

Darrel Moellendorf’s recent work on climate change and intergenerational justice is particularly relevant to this discussion because he implicitly responds to the challenge of Gardiner’s PIP (2009). Moellendorf defends a principle of Intergenerational Equality that states, “Present energy policy should produce foreseeable future (adaptation) costs of CO<sub>2</sub> emissions whose proportion to overall future economic output is equal to the proportion of (mitigation) costs to output of the present generation” (2009, p. 207). His argument appeals to a contractualist view of procedural justice that draws on John Rawls’s original position argument (Rawls 2001). Moellendorf’s key claim is that in order to achieve impartiality in our deliberations about intergenerational justice, we must bind ourselves “only to principles that we would find it acceptable for previous generations to have bound themselves” (2009, p 211). Moellendorf believes this may alleviate Gardiner’s worries about the implications of PIP insofar as committing to procedural justice may enable us to make moral commitments that extend us beyond a generationally self-interested perspective. (Gardiner himself, however, identifies several problems with contractualist views in intergenerational settings (2009).) With respect to intergenerational risks, Moellendorf’s view suggests we ought to take a perspective that will be fair across generations with respect to the costs of climate change. Of note, however, is that Moellendorf includes a “foreseeability” clause that limits the principle of Intergenerational Equality to known (or knowable) climate impacts. This implies

that policies condoning greenhouse gas emissions before the consequences of such emissions were known were not unfair or unjust. In order to be unjust on this view, a risky activity must be understood as such.

Another perspective comes from Henry Shue, who has written extensively on intergenerational justice with an emphasis on climate change and other environmental issues (e.g., 1980, 1981, 1993, 1999, 2010). His work on basic rights has influenced many writing on climate justice in particular. The intuition grounding his work on basic rights is that “human beings ought to enjoy control over avoidable damage to their own bodies” (Shue 1981, p. 593). Basic rights then, “are the morality of the depths. They specify the line beneath which no one is to be allowed to sink” (Shue 1996, p. 18). Basic rights specify “everyone’s minimum reasonable demands upon the rest of humanity” (Shue 1996, p. 19). Implicitly drawing on his work on basic rights, Shue (1993) makes an influential distinction between subsistence emissions and luxury emissions. Shue argues that we have to ensure the poor are entitled to those emissions that enable them to meet their basic needs, namely, to subsistence emissions. Part of what this implies is that if there is to be a system of emission allowances, justice requires that we not allow the wealthy to buy up the poor’s emission allowances. (In his 1999 paper, Shue similarly makes a multi-pronged argument, drawing on three distinct principles of justice, that wealthy industrialized states that should at least initially bear the costs of addressing global environmental problems.) We can apply this to the present discussion of intergenerational risks by understanding Shue as suggesting that whatever our obligations to minimize the extent of intergenerational risks are, we must ensure people’s basic needs are met. Extending his view may suggest that our priority in addressing intergenerational risks is to ensure both present and future people’s basic needs are met. Risky activities that threaten present or future people’s ability to satisfy their basic needs are morally unacceptable on this view.

Simon Caney draws heavily on Shue’s work in defending a human rights-based approach to climate change policy (2010a, b). Caney proposes that we understand dangerous climate change as climate change that systematically undermines the widespread enjoyment of human rights. He understands human rights as representing moral thresholds that have, to use John Rawls’ terminology, lexical priority over other moral values (Caney 2010a; see Rawls 1999, pp. 37–38 for a discussion of “lexical priority”). Caney concentrates on three key human rights: the human rights to life, health, and subsistence, focusing on what he takes to be the least contentious formulations of these rights to show that on even the most minimal conception of human rights it is clear that climate change jeopardizes these rights. He phrases each of these rights in such a way that individuals have a right not to have their ability or access to life, health and subsistence taken away or interfered with by other people. Caney thus frames human rights in terms of what people owe to each other, which allows him to conclude that whoever is contributing to climate change is violating the rights of individuals whose health, subsistence, and/or existence is threatened by the effects of climate change. We may apply Caney’s view to intergenerational risks when we conceive of some such risks as threatening to violate future people’s human rights. On this view we ought not create or contribute to intergenerational risks that threaten present or future people’s rights to life, health, and subsistence.

Caney defends two principles, which are revised versions of principles first defended by Shue (1999), as a starting point for thinking about climate change policy. These are, first the Poverty-Sensitive Polluter Pays Principle, which says that whoever contributes to dangerous climate change should bear the burden of combating it, unless this would push him/her below a decent minimal standard of living. And second, the History-Sensitive Ability to Pay Principle

which says that any remaining burden (not covered by the first principle – e.g., past emissions, nonhuman induced emissions, and legitimate emissions of the poor) should be borne by the wealthy, proportionally to their wealth and to whether they acquired this wealth in climate endangering or other unjust ways (Caney 2010b). One important implication of Caney's "hybrid model" that he points out is that the least advantaged have a duty to develop in non-carbon-intensive ways, if they can do so without great cost to themselves (2010b). (Caney admits that his view does not specify the appropriate duty-bearing entities, which would be necessary in order to formulate climate change policy. He notes that most discussions focus primarily on states as the bearers of responsibility in climate change policy, but points out that this focus might ignore important differences between people within states and other entities such as firms.) More generally, these principles may be applied to intergenerational risks that threaten human rights such that those contributing to intergenerational risks ought to combat potential human rights violations and if this is insufficient the wealthy ought to help protect future people's human rights.

Steve Vanderheiden (2008) defends a different view of what justice requires of us in climate change policy. His most distinctive move is to differentiate what justice requires for mitigation vs. adaptation policy because, he argues, these represent different distributive issues. Vanderheiden argues that mitigation policy, or emissions abatement policy, as he describes it, should be guided by an equity model. This model assumes that all people are moral equals and therefore have equal claims to the atmosphere's absorptive capacities, suggesting a default position of equal per-capita emissions shares. An unequal distribution of the atmosphere's absorptive capacity would have to be justified based on benefits accrued to the least advantaged. Adaptation and compensation policy, on the other hand, should be guided, argues Vanderheiden, by considerations of responsibility (for contributing to climate change). So while mitigation policy should look forward toward equal distributions of emissions, adaptation policy should look backward toward liability for creating threats of harm.

Vanderheiden, also following Shue, distinguishes between survival and luxury emissions such that he argues people have a basic right to survival (but not luxury) emissions. Thus while he too takes a rights-based approach in at least this sense, he comes to much different conclusions about what this implies than Caney. Vanderheiden brings rights into his account in part to highlight that the right to survival emissions is but one such right and that there is also a right to develop. He stresses the importance of thinking about the relationship between climate change and development policies. This implies that Vanderheiden may argue that how we address intergenerational risks cannot be separated from other justice issues. Yet he does not seem to address or acknowledge the contributions to climate change that pose intergenerational risks. On the other hand, failing to respect the right to develop could be interpreted as posing intergenerational risks such that Vanderheiden is actually taking a wide view of intergenerational risks.

What this very brief discussion of climate justice is meant to illustrate is that work on intergenerational justice has a lot to offer toward understanding the ethics of intergenerational risks. Caney and Vanderheiden's theories, for example, illustrate the way in which theories of climate justice (or intergenerational justice more generally) can guide decision making about intergenerational risks. Often, though not always, we can frame the question of whether or not a risky activity or behavior that will have long-term consequences (and hence poses an intergenerational risk) is morally acceptable or not as a question about whether such an activity or behavior is unjust or not. Not all of the ethical issues raised by intergenerational risks may turn out to be justice issues, but intergenerational risks certainly pose at least some justice questions.

## Further Research

---

That the very meaning of “intergenerational risks” requires expanding our existing concepts is illustrative of the complexity of such risks. Once we realize the far-reaching implications of our actions and we recognize that many risky activities will affect who will live in the future, we are forced to recognize that intergenerational risks may be more common and pervasive than we may have at first thought. And once we further realize the temptations of ignoring or avoiding the ethical issues intergenerational risks present, as Gardiner highlights, that intergenerational risks merit much greater attention, study, and consideration becomes clear. There are existing bodies of literature that inform how we should think about and address intergenerational risks, but we need to be much more explicit about this task if we are to do so effectively and appropriately. Considering how much, if any, knowledge an actor must have of the potentially harmful nature of her actions in order to be held responsible for the risky nature of such actions, for example, is an important consideration that merits greater attention than I (and some of the authors I have considered) have given it here. Much more work needs to be done in intergenerational ethics in order for us to be able to understand how we should think about the challenges intergenerational risks pose and how we should address them. What we think we owe future people, after all, most certainly will affect what (intergenerationally) risky activities we deem to be morally acceptable.

Another important issue that I have yet to so much as touch on is that of nonanthropocentric ethics. So far the entire discussion has assumed that intergenerational risks threaten future humans with no reference to the potential noninstrumental value of the nonhuman world. But we can imagine some risks as posing threats of harm to nonhuman animals and/or nature if we expand our interpretation of what constitutes “harm.” There are many theories in environmental ethics that attribute value to various aspects of the nonhuman world (which may or may not conflict with anthropocentric values) and which would worry about activities or events that threaten the future integrity of natural ecosystems, species, and/or ability of particular organisms to thrive and survive. As discussed here, threats to the environment may be bad because of the instrumental value of the nonhuman world, but many environmentalists and environmental ethicists would see deeper wrongs in environmental degradation or damage. It may be that we should understand long-term risks (as opposed to intergenerational risks) as *de dicto* threats of harm to the environment; such an interpretation would very much change how we should address such risks. There is a much larger, highly relevant debate about what has (intrinsic) moral value that is relevant to how we think about long-term risks. One’s moral perspective will inform one’s understanding of how we should think about and address intergenerational (or maybe more appropriately long-term) risks.

The lack of much direct work on the ethics of intergenerational risks should push us to question our most basic assumptions about how we should think about risks and the fact that so many risks are in fact intergenerational risks. I hope that understanding intergenerational risks will expose both the prevalence and moral complexity of such cases. This in turn will hopefully reveal the need for more thorough and deliberate contemplation of how we should think about intergenerational ethics. Future work should aim to better understand the nature of intergenerational risks and how these differ from intragenerational risks. This will hopefully enable us to more thoroughly consider what is at stake when we engage in intergenerational risky activities and to assess what principles and/or approaches should guide our actions that may create or contribute to intergenerational risks, which pose threats of *de dicto* harm on future generations.

## Acknowledgment

---

I began thinking about intergenerational risks during my undergraduate studies at Connecticut College. I am indebted to Derek Turner for his work with me on my undergraduate thesis as well as his continued support, guidance, and comments on an earlier draft of this chapter. I am also indebted to Stephen Gardiner who made this chapter possible and whose insightful comments added depth to my discussion. My graduate dissertation committee, Debra Satz, Joshua Cohen, Tamar Schapiro, and Stephen Schneider, also provided me with guidance and support that greatly influenced my work in this chapter. I am grateful for having had such amazing and inspiring mentors.

## References

---

- Broome J (2006) Valuing policies in response to climate change: some ethical issues. A contribution to the work of the Stern review of the economics of climate change. Discussion Paper, University of Oxford
- Caney S (2008) Human rights, climate change, and discounting. *Environ Polit* 17:536–555
- Caney S (2009) Climate change and the future: discounting for time, wealth, and risk. *J Soc Philos* 40:163–186
- Caney S (2010a) Climate change, human rights and moral thresholds. In: Humphreys S (ed) *Human rights and climate change*. Cambridge University Press, Cambridge, pp 69–90
- Caney S (2010b) Climate change and the duties of the advantaged. *Crit Rev Int Soc Polit Philos* 13: 203–228
- Dietz S, Hope C, Sern N, Zenghelis D (2007) A robust case for strong action to reduce the risks of climate change. *World Econ* 8:121–168
- Feinberg J (1990) *Harmless wrongdoing: the moral limits of the criminal law*. Oxford University Press, New York
- Gardiner S (2002) The real tragedy of the commons. *Philos Public Aff* 30:387–416
- Gardiner S (2003) The pure intergenerational problem. *The Monist* 86:481–500
- Gardiner S (2006a) A perfect moral storm: climate change, intergenerational ethics and the problem of moral corruption. *Environ Values* 15:397–413
- Gardiner S (2006b) A core precautionary principle. *J Polit Philos* 14:33–60
- Gardiner S (2009) *A contract on future generations?* In: Gosseries A, Meyer LH (eds) *Intergenerational justice*. Oxford University Press, Oxford, pp 77–118
- Gardiner S (2011) *A perfect moral storm: the ethical tragedy of climate change*. Oxford University Press, Oxford
- Goklany I (2001) *The precautionary principle: a critical appraisal of environmental risk assessment*. Cato Institute, Washington, DC
- Gosseries A, Meyer L (eds) (2009) *Intergenerational justice*. Oxford University Press, Oxford
- Hansson S (2007) Risk and ethics: three approaches. In: Lewens T (ed) *Risk: philosophical perspectives*. Routledge, London, pp 21–35
- Hare C (2007) Voices from another world: must we respect the interests of people who do not, and will never, exist? *Ethics* 117:498–523
- Harman E (2003) Moral status. Dissertation submitted to Massachusetts Institute of Technology
- Harman E (2004) Can we harm and benefit in creating? *Philos Perspect* 18:89–113
- Harman E (2009) Harming as causing harm. In: Roberts M, Wasserman D (eds) *Harming future persons: ethics, genetics and the nonidentity problem*. Springer, New York, pp 137–154
- Hartzell L (2009) Rethinking the precautionary principle and its role in climate change policy. Dissertation submitted to Stanford University
- Hickey J, Walter V (1995) Refining the precautionary principle in international environmental law. *Va Environ Law J* 14:423–454
- IPCC, Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) (2007) *Climate change 2007: the physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge University Press, Cambridge/New York, 996 pp
- Kight F (2002) *Risk, uncertainty and profit*. Beard Books, Washington, DC
- Kuman R (2003) Who can be wronged? *Philos Public Aff* 31:99–118
- Kusch M (2007) Towards a political philosophy of risk: experts and publics in deliberative democracy.

- In: Lewens T (ed) *Risk: philosophical perspectives*. Routledge, London, pp 131–155
- Lewens T (ed) (2007) *Risk: philosophical perspectives*. Routledge, London
- Manson N (2002) Formulating the precautionary principle. *Environ Ethics* 24:263–274
- Moellendorf D (2009) Justice and the assignment of the intergenerational costs of climate change. *J Soc Philos* 40:204–224
- Neumeyer E (2007) A missed opportunity: the Stern review on climate change fails to tackle the issue of non-substitutable loss of natural capital. *Global Environ Change* 17:297–301
- Nordhaus W (2007a) A review of the Stern Review on the economics of climate change. *J Econ Lit* 45:686–702
- Nordhaus W (2007b) Critical assumptions in the Stern review on climate change. *Science* 317:201–202
- Nordhaus W (2008) A question of balance: weighing the options on global warming policies. Yale University Press, New Haven
- O'Riordan T, Jordan A (1995) The precautionary principle in contemporary environmental politics. *Environ Values* 4:191–212
- Page E (2006) Climate change, justice and future generations. Edward Elgar, Northampton
- Parfit D (1976) On doing the best for our children. In: Bayles M (ed) *Ethics and population*. Schenkman, Cambridge, pp 100–115
- Parfit D (1984) Reasons and persons. Clarendon, Oxford
- Parfit D (2001) Energy policy and the further future: the identity problem. In: Pojman L (ed) *Environmental ethics: readings in theory and application*, 3rd edn. Wadsworth, Stamford, pp 289–296. Originally published in: MacLean D, Brown P (eds) *Energy and the future*. Rowman & Littlefield, Totowa, pp 166–179
- Perry S (2007) Risk, harm, interests, and rights. In: Lewens T (ed) *Risk: philosophical perspectives*. Routledge, London, pp 190–209
- Posner R (2001) Cost-benefit analysis: definition, justification, and comment on conference papers. In: Adler M, Posner E (eds) *Cost-benefit analysis: legal, economic, and philosophical perspectives*. University of Chicago Press, Chicago, pp 317–341
- Posner R (2004) *Catastrophe: risk and response*. Oxford University Press, Oxford
- Raffensperger C, Tickner J (1999) Introduction: to foresee and to forestall. In: Raffensperger C, Tickner J (eds) *Protecting public health and the environment: implementing the precautionary principle*. Island Press, Washington, DC, pp 1–11
- Rawls J (1999) *A theory of justice*, revised edn. Harvard University Press, Cambridge
- Rawls J (2001) *Justice as fairness*. The Belknap Press of Harvard University Press, Cambridge
- Reiman J (2007) Being fair to future people: the non-identity problem in the original position. *Philos Public Aff* 35:69–92
- Resnik D (2003) Is the precautionary principle unscientific. *Stud Hist Philos Biol Biomed Sci* 34:329–344
- Roberts M, Wasserman D (eds) (2009) *Harming future persons: ethics, genetics and the nonidentity problem*. Springer, New York
- Sandin P (2004) The precautionary principle and the concept of precaution. *Environ Values* 13:461–465
- Sandin P (2007) Common-sense precaution and varieties of the precautionary principle. In: Lewens T (ed) *Risk: philosophical perspectives*. Routledge, London, pp 99–112
- Sen A (1995) Environmental evaluation and social choice: contingent valuation and the market analogy. *Jpn Econ Rev* 46:23–37
- Shue H (1980) *Basic rights: subsistence, affluence, and U.S. foreign policy*. Princeton University Press, Princeton
- Shue H (1981) Exporting hazards. *Ethics* 91:579–606
- Shue H (1993) Subsistence emissions and luxury emissions. *Law & Policy* 15:39–59
- Shue H (1996) *Basic rights: subsistence, affluence, and U.S. foreign policy*, 2nd edn. Princeton University Press, Princeton
- Shue H (1999) Global environment and international inequality. *Int Aff* 73:531–545
- Shue H (2010) Deadly delays, saving opportunities: creating a more dangerous world? In: Gardiner M, Caney S, Jamieson D, Shue S (eds) *Climate ethics: essential readings*. Oxford University Press, Oxford, pp 146–162
- Sinnott-Armstrong W (2005) It's not *my* fault: global warming and individual moral obligations. In: Sinnott-Armstrong W, Howarth R (eds) *Perspectives on climate change*. Elsevier, Amsterdam, pp 221–253
- Stern N (2007) The economics of climate change: the stern review. Cambridge University Press, Cambridge
- Sunstein C (2005) *Laws of fear: beyond the precautionary principle*. Cambridge University Press, Cambridge, MA
- Tickner J (2003a) The role of environmental science in precautionary decision making. In: Tickner J (ed) *Precaution, environmental science and preventive public policy*. Island Press, Washington, DC, pp 3–19
- Tickner J (2003b) Precautionary assessment: a framework for integrating science, uncertainty, and preventive public policy. In: Tickner J (ed) *Precaution, environmental science and preventive public policy*. Island Press, Washington, DC, pp 265–278
- Turner D, Hartzell L (2004) The lack of clarity in the precautionary principle. *Environ Values* 13:449–460

- Vanderheiden S (2008) Atmospheric justice: a political theory of climate change. Oxford University Press, Oxford
- Weitzman M (2007) A review of the Stern report on the economics of climate change. *J Econ Lit* 45:703–724
- Whiteside K (2006) Precautionary politics: principle and practice in confronting environmental risk. MIT Press, Cambridge
- Woodward J (1986) The non-identity problem. *Ethics* 96:804–831

# 38 The Precautionary Principle

Marko Ahteesuu<sup>1</sup> · Per Sandin<sup>2</sup>

<sup>1</sup>University of Turku, Turku, Finland

<sup>2</sup>Swedish University of Agricultural Sciences, Uppsala, Sweden

<b>Introduction .....</b>	<b>962</b>
<b>History .....</b>	<b>963</b>
General Idea of Precaution .....	963
Specific Codes of Conduct and Arguments from Precaution .....	965
Judicial Documents .....	966
<b>Current Research .....</b>	<b>967</b>
Terminology .....	967
Basic Structure .....	969
The Weak/Strong Distinction .....	970
Types of Precautionary Principles .....	971
Arguments Against the Precautionary Principle .....	972
<b>Further Research .....</b>	<b>973</b>
Formal Methods .....	973
Normative Underpinnings .....	974
Risk Analysis .....	974
Participatory Decision-Making Practices .....	975

**Abstract:** The precautionary principle has come to the fore in risk discourse. It calls for early measures to avoid and mitigate environmental damage and health hazards in the face of uncertainty. This paper reviews the history of and current research on the principle and suggests areas where further scrutiny is needed. The origin of the precautionary principle is traced back to three sources: (1) the general idea of precaution, (2) specific nonjudicial codes of conduct and arguments from precaution, and (3) law texts. Much of the current theoretical study has been concerned with analysis of different aspects of the precautionary principle and with assessment of specific versions of the principle in different regulatory contexts. Issues related to terminology, to basic structure shared by different formulations of the principle, and to the distinction between strong and weak interpretations are considered. Discussion on different versions of the precautionary principle as (1) rules of choice, (2) procedural requirements, and (3) epistemic rules or principles is briefly revisited. General arguments leveled at the principle are spelled out. Despite the attention the precautionary principle has received in academic literature, there remain areas of research which deserve more thorough scrutiny. Formal methods of inquiry have been insufficiently utilized. Topics which deserve further study include the normative underpinnings of the principle, the status of the principle in scientific risk analysis, and the principle's relationship with stakeholder/public engagement.

## Introduction

---

The precautionary principle embodies the idea that *in dubio pro natura*. If in doubt, decide in favor of the environment. Over the past two and half decades, this seemingly simple precept has found its way to the center of environmental law and policy. The precautionary principle is mentioned in domestic laws (e.g., in Australia, France, Germany, Sweden, and UK) and international treaties (e.g., CPB 2000). Numerous declarations and other soft law instruments include the principle in their objectives or articles/general principles (e.g., UNCED 1992). Common examples come from the contexts of marine protection and disposal of hazardous wastes, fisheries management, protection of the ozone layer and chemicals regulation, conservation of the natural environment and biological diversity, climate change and global warming policies, regulation of genetically modified organisms (GMOs), and public health policy.

Arguably, the most noted formulation of the precautionary principle is the one adopted at the *United Nations Conference on Environment and Development* in Rio de Janeiro.

- ▶ Where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation (UNCED 1992, Principle 15).

Another well-known formulation was introduced at a conference organized by the Science and Environment Health Network (SEHN) in 1998. According to it,

- ▶ [w]hen an activity raises threats of harm to human health or the environment, precautionary measures should be taken even if some cause and effect relationships are not fully established scientifically (*Wingspread Statement on the Precautionary Principle* 1998).

Besides the Rio formulation and the *Wingspread Statement*, many other formulations of the precautionary principle can be found in laws, treaties, protocols, declarations, communications, and other legal and policy documents (for legal analyses of the principle see Fitzmaurice 2009;

Hohmann 1994; de Sadeleer 2002; Trouwborst 2002). Several definitions or analyses have been put forth by academic scholars in the relevant theoretical literature (see Adams 2002; Gardiner 2006; Manson 2002; Morris 2000; Sandin 1999, 2004b; von Schomberg 2006; Soule 2002; Sunstein 2005).

Not only does the precautionary principle take many forms, but it remains a matter of ongoing controversy. Academic scholars from various disciplines (e.g., risk analysts, legal theorists, economists, political scientists, and ethicists), decision-makers, industry representatives, environmental organizations, and the lay people continuously argue about the principle. The debate is centered mainly on two interrelated issues. First, despite academic efforts to clarify the principle and established policy documents, consensus has not been reached concerning the exact definition of the principle. Second, the way in which the precautionary principle should be put into practice has remained controversial. No commonly accepted guidelines for the implementation of the principle exist. This holds even in the European Union (EU), where the precautionary principle has gained the most attention and popularity. In spite of the *Communication on the Precautionary Principle* (CEC 2000), which was introduced by the Commission of the European Communities in order to standardize the use of the principle, the adopted national precautionary policies within the EU have varied in a wide range (see e.g., Levidow et al. 2005).

Furthermore, the precautionary principle has been criticized by some scholars such as political scientist Aaron Wildavsky (1996), legal scholar Cass R. Sunstein (2005), economist Julian Morris (2000), physicist Chauncey Starr (2003), and bioethicists John Harris and Søren Holm (e.g., Holm and Harris 1999; Harris and Holm 2002). Most commonly, the principle is claimed to be too vague, incoherent, unscientific, and/or counterproductive. Theoretical differences are deep-seated. Some authors do not even seem to consider the precautionary principle an action-guiding precept, but purely an epistemic (i.e., belief-guiding) or a procedural principle. Although the precautionary principle is predominantly a legal principle, it has been regarded as a decision rule (Hansson 1997), epistemic principle (Peterson 2007), a risk management tool (CEC 2000, pp. 3, 13), an ethical principle (Carr 2002), a methodological rule for risk research (Tickner 2003), and an organizing concept for various contemporary ideas that challenge the regulatory *status quo* (Jordan and O'Riordan 1999; for discussion see Gardiner 2006; Harris and Holm 2002; Sandin 2007).

In what follows, we will review the history of and current research on the precautionary principle and suggests areas where further scrutiny is needed.

## History

Various views have been proposed concerning the origin of the precautionary principle. There is also a disagreement over the first appearance of the principle in official documents (see e.g., Adams 2002). Different accounts that trace back the origin of the precautionary principle may be subsumed under three classes. These include (1) the general idea of precaution, (2) specific nonjudicial codes of conduct and arguments from precaution, and (3) official documents.

### General Idea of Precaution

It has been argued that the origin of the precautionary principle can be found in the general and everyday idea of precaution. Philippe H. Martin from the European Commission Joint Research Centre claims that

- ▶ [t]he precautionary principle is an age-old concept. Unambiguous reference to precaution as a management guideline is found in the millennial oral tradition of Indigenous People of Eurasia, Africa, the Americas, Oceania, and Australia (Martin 1997, p. 276).

This view reflects a very wide understanding of the precautionary principle not restricted to environmental and health concerns. Precaution has played a role in oral traditions around the globe. It guides us not to inflict harm with our actions. The core of the principle is seen as a rule of thumb.

Taking precautions is no doubt in accordance with common sense. “The idea that care and foresight are required in the face of (...) uncertain future is universal and of all times” (Trouwborst 2002, p. 7). As a number of scholars have noted, the essence of precaution is captured by several English sayings such as “better safe than sorry,” “look before you leap,” “a stitch in time saves nine,” and “an ounce of prevention is worth a pound of cure” (e.g., Randall 2011; Resnik 2003; Sandin 2004b; Trouwborst 2002; VanderZwaag 2002). Despite the apparent similarities, it should however be noted that these sayings and aphorisms are general in nature. They do not provide specific guidance for concrete situations. The sayings emphasize the avoidance of harm and preparing for the uncertain future, but they seem to do it in a wider sense than does the precautionary principle. Unlike the general idea of precaution, the principle is triggered only by inadequate and/or disputed scientific knowledge concerning certain types of environmental threats and health hazards.

Even if we sidestep the obvious differences in content between precaution and the precautionary principle, we might still end up with a more theoretical problem. Even if someone takes precautions in a particular situation, this does not necessarily mean that the agent in question subscribes to a precautionary principle. In order to infer that there is a precautionary principle present, we should at least demand that the agent holds that precaution should be taken in the particular situation and in situations that are relevantly similar, with criteria for relevant similarity further specified in some way (Sandin 2004a, p. 4). This minimum condition is not fulfilled in the case of sayings or the other writings mentioned above because the similar cases in which precaution should be applied are not specified.

Instead of regarding precaution to be equivalent to the precautionary principle, it can still be considered an origin or a predecessor. The general idea of precaution might also be used to explain the remarkable attention that the precautionary principle has received. According to legal scholar David VanderZwaag (2002, p. 166), “[a] prime reason for the international popularity of precaution is its reflection of common sense notions evident in numerous cultures.” The principle provides a practical implementation to several wisdoms or sayings. VanderZwaag is right in that part of the evident “magnetism” of the precautionary principle springs from the intuitiveness of the general idea of precaution. Yet, other factors may have played a greater role.

First, owing to a number of factors (such as the growth in the world’s population; the increasing change, complexity, and interdependencies of societies; and the new possibilities provided by rapid technological development), the stakes have become higher than before. Human action can lead – and has already contributed – to serious and irreversible environmental damage. We may be facing a changed situation with regard to the inducement and management of environmental threats and health hazards. Second, a growing recognition of ecosystems’ sensitivity as well as of their intra- and interdependencies is not without significance. Our limited understanding of several natural processes and related risks has increasingly

been admitted and emphasized. Third, the prevailing institutionalized risk governance methodology (especially quantitative risk assessment) has been subjected to criticism. In particular, the strict divisions between science and value and between risk assessment and risk management have been called into question.

In sum, various forms of precaution have been taken as long as human beings have existed. It would however be problematic to argue that the precautionary principle has actually been invoked in the above-mentioned cases. The general idea of precaution has most probably played a role in the formation of the precautionary principle, and it can be employed to partly explain the wide endorsement of the principle. (For an analysis of the concept “taking precautions” see Sandin 2004b).

### Specific Codes of Conduct and Arguments from Precaution

---

The origin of the precautionary principle has sometimes been traced to specific nonjudicial codes of conduct and arguments from precaution. According to the former view, the basic idea of the principle has been present in ethical codes and policies, e.g., in public health policy. Some scholars have even argued that the first reference to it can be found in the Hippocratic Oath *primum non nocere* (first, do no harm) (Ozonoff 1999, p. 100; Graham and Hsia 2002, p. 374). We agree that certain ethical codes and policies might be correctly attributed as being precautionary in nature. Nevertheless, these codes and policies typically seem to resemble more closely the general idea of precaution than the precautionary principle itself (as the latter is usually understood).

The second strategy is to trace the history of the precautionary principle to certain arguments from precaution which have been presented in various contexts, e.g., in energy policy criticism. This is the case when Morris asserts that

- ▶ PP-like arguments have been used in the USA since the 1950s; at that time, groups of political conservatives opposed fluoridation of water on the grounds that fluoride was used as rat poison and that involuntary fluoridation amounted to mass medication, a step on the road to socialism. In the 1960s, left-wing radicals similarly used PP-like arguments against nuclear power. (...) By highlighting th[e] possibility of catastrophe, regardless of the probability of its occurrence, campaigners were able to instill fear of the technology as such (Morris 2000, p. 2; see also Goodin 1980, pp. 418–419; Pearce et al. 1980, p. 58).

This strategy is also put forward by philosophers Derek Turner and Lauren Hartzell. They contend that German philosopher Hans Jonas gave an early version of the precautionary principle in his book *The Imperative of Responsibility* (Turner and Hartzell 2004, p. 452). In Jonas' words, “[i]t is the rule, stated primitively, that the prophecy of doom is to be given greater heed than the prophecy of bliss” (Jonas 1984, p. 31). Although Jonas' vision is general in nature, it has clear similarity to the very basic idea of the precautionary principle. First, the principle implies that when considering the introduction of an activity, certain environmental and health risks outweigh the possible (economic) benefits of that activity. Second, Jonas is concerned about the possible irrevocable consequences of technological developments, and this also reflects the common understanding of the principle.

In addition to the cited examples, there have been several other claims about the origin of the precautionary principle which can be subsumed under this class (see, e.g., EEA 2001;

Martin 1997, p. 264). Given the apparent similarities and lines of development of ideas, these arguments from precaution deserve to be notified when the history of the precautionary principle is under study.

## Judicial Documents

---

The most common strategy to search for the origin of the precautionary principle is to look for instances of it in law texts and policy documents. This may take two forms. The first option is to identify early explicit use of the term “precautionary principle” or other equivalent terms. The problem with this strategy arises from the facts that the first references to the principle were often brief and that the principle was not defined. Moreover, several phrases have been employed, such as the terms “precautionary measures” and “precautionary approach.”

A commonly agreed predecessor of the precautionary principle is the *Vorsorgeprinzip* which was introduced to German environmental law and policy in the 1970s (for discussion see Boehmer-Christiansen 1994). The *Vorsorgeprinzip* emphasizes identification of early warnings of environmental threats and preparing beforehand for the uncertain future and for its risks. It was first incorporated into air and water protection act in West Germany, but it soon became a fundamental principle of German environmental law.

The first explicit mention of the precautionary principle (or more precisely, a precautionary approach) in an international environmental declaration was in 1987. The Ministerial Declaration of the *Second International Conference on the Protection of the North Sea* states that

- ▶ [a]ccepting that in order to protect the North Sea from possibly damaging effects of the most dangerous substances, a *precautionary approach* is necessary which may require action to control inputs of such substances even before a causal link has been established by absolutely clear scientific evidence (Paragraph 7; Italics added).

The second option to search for the origin of the precautionary principle is to identify cases in which the principle is thought to be present even if it has not been explicitly mentioned. Accordingly, the first instance in an official text may be found in the *World Charter for Nature*, adopted by the United Nations General Assembly, as early as in 1982.

- ▶ Activities that are likely to pose a significant risk to nature shall be preceded by an exhaustive examination; their proponents shall demonstrate that expected benefits outweigh potential damage to nature, and where potential adverse effects are not fully understood, the activities should not proceed (Principle 11b; Italics added).

Principle 11b includes the basic constituents of the precautionary principle: a reference to an unacceptable threat of environmental damage and/or health hazard, to scientific uncertainty, and to precautionary measures in the form of inactions.

Other cases have also been suggested. Legal scholar Daniel Bodansky (1991) argues for the early use of the precautionary principle in law and policy. In his view, although the principle is not explicitly mentioned in the environmental law of the USA, the precautionary principle has been the basis of much of it for several years. The “no-discharge” requirement of the 1972 *Federal Water Pollution Control Act Amendments* is provided as an example of this. The act

presumed that discharges of pollutants are harmful to the water quality in the first place even without any scientific predictions, and it included the so-called ALARA (as low as reasonably achievable) approach as the required response. Early examples of the precautionary principle can also be found at international level. In the 1970s, a moratorium on commercial whaling “was justified on the basis of uncertainty about the impacts of continued whaling (...) rather than on the basis of scientific evidence” (Bodansky 1991, p. 5). According to Mikael Karlsson, president of European Environmental Bureau (EEB), the precautionary principle was included in the *Swedish Environmental Protection Act* of 1969 (Karlsson 2006; for other examples of this strategy, see also EEA 2001; Myers 2006, pp. 4–6; Whiteside 2006, pp. 66–70).

The further development of the precautionary principle in environmental law and policy is not reviewed here. Several such analyses are available (Fitzmaurice 2009; Freestone and Hey 1996; O’Riordan et al. 2002; de Sadeleer 2006; de Sadeleer 2002; Trouwborst 2002; for a comparison of precaution in the USA and Europe see Whiteside 2006, Chap. 3; Wiener and Rogers 2002).

## Current Research

---

Much of the current theoretical research on the precautionary principle has been concerned with clarification of its different aspects, with analysis of its relevance in different regulatory contexts, and with arguments for and against its different interpretations. Next, discussion on terminological issues, basic structure, weak/strong distinction, three types of the precautionary principle, and main arguments against the principle will be briefly revisited.

## Terminology

---

Terminological issues present a source of confusion and disagreement. Most authors speak about one definite principle (e.g., Rogers 2001), but others use the indefinite plural form (e.g., Löfstedt et al. 2002). Furthermore, several terms have been employed. In official documents, phrases such as “precautionary measures,” “precautionary principle,” “principle of precautionary action,” and “precautionary approach” can be found. An even more diverse set of phrases has been used in the commentary literature pertaining to the principle. Besides the aforementioned terms, the examples include “precaution” (Levidow et al. 2005), “precautionary thinking” (Trouwborst 2002), “precautionary assessment” (Tickner 2003), “precautionary science” (Cranor 2003), and “precautionary principle/approach” (see e.g., VanderZwaag 2002; cf. Peterson 2007).

The following points make it understandable that some legal scholars (e.g., Hohmann 1994) use the plural form, i.e., *precautionary principles*, when they refer to the precautionary principle. As noted, there are several formulations of the principle in official documents. The formulations differ from each other. Although differences are to be expected between different regulatory contexts given the different situations in regard to risk imposition, knowledge about risks and manageability of risks, formulations differ considerably also within a regulatory context. Consider an example of this found in marine environmental protection. The Ministerial Declaration of the *Second International Conference on the Protection of the North Sea states*

that precautions can be taken in the absence of absolutely clear scientific evidence. In the corresponding statement of the third conference, no evidence to prove the causal connection is required.

- ▶ [A] precautionary approach is necessary which may require action to control inputs of such substances even before a causal link has been established by absolutely clear scientific evidence (*Second International Conference on the Protection of the North Sea 1987*, Paragraph 7).
- ▶ [P]recautionary principle, that is to take action to avoid potentially damaging impacts of substances that are persistent, toxic and liable to bioaccumulate even when there is no scientific evidence to prove a causal link between emissions and effects (*Third International Conference on the Protection of the North Sea 1990*, Preamble).

Bodansky (1991, p. 5) has phrased an obvious worry lying in the background: because different formulations of the precautionary principle state the trigger for taking precautions and the appropriate precautionary measures in radically differing ways, “it is difficult to speak of a single principle at all.” Nevertheless, instead of speaking about several principles, (at least, apart from judicial studies) it seems to be more fruitful to say that there is only one principle which is formulated (or understood) in various ways. The use of singular and plural may just indicate the fact that the precautionary principle is thought of at different levels of generality.

In the academic literature, different terms have been employed to point out slight differences between theoretical positions. As an example of this, environmental health scholar Joel Tickner (2003) has introduced precautionary assessment in which the principle has implications for the risk assessment phase. This goes against the typical view of the precautionary principle as merely a risk management tool. That one concept has sometimes been used in several meanings is also worth noticing. The term “precaution” has been employed to refer to the precautionary principle (e.g., Levidow et al. 2005), to the prescribed precautionary measures, and to the general idea of precaution (Sandin 2004b).

The most debated terminological issue is the possible disparity between the terms “precautionary principle” and “precautionary approach.” It is not straightforwardly clear as to whether there is a difference in meaning between the precautionary principle and the precautionary approach (see e.g., Conko 2003, pp. 642–643; Trouwborst 2002, pp. 3–5; VanderZwaag 2002, pp. 166–167). It has been thought that the precautionary approach represents a less stringent version of the precautionary principle and that it thus avoids theoretically implausible forms of absolutism such as total reversal of the burden of proof (see e.g., Conko 2003, p. 642). One context in which the terms are commonly distinguished from each other is fisheries management (see e.g., Mace and Sissenwine 2002, pp. 14–15; Ortega Vicuña 1999). Several authors nevertheless hold the opposite view that the terms “precautionary principle” and “precautionary approach” can be used interchangeably (e.g., Mascher 1997, p. 70; Trouwborst 2002, p. 4). It is noteworthy that some official documents treat the terms as equivalents (e.g., *Wingspread Statement 1998*). The formulation of Principle 15 of the Rio Declaration (UNCED 1992) which is typically considered a paradigm example of the precautionary principle actually includes the term “precautionary approach” in the English versions of the text, while translations into several other languages (e.g., Swedish) use terms corresponding to “precautionary principle.”

We employ the term “precautionary principle” as a uniting term for the various phrases found in official documents and in the relevant academic commentary literature.

The disparities that the use of different terms sometimes implies are taken into consideration by means of speaking about different understandings (or interpretations/formulations) of the principle.

## Basic Structure

---

Many scholars have argued that (several) different formulations of the precautionary principle share common elements and a common structure. It has been spelled out in slightly different ways and terms by different authors adopting what might be called an analytical approach to the precautionary principle (e.g., Sandin 1999; Manson 2002). These structural schemata can be used to explicate, compare, and evaluate different formulations of the precautionary principle. An early attempt is Sandin who considers the dimensions of the principle and provides the following characterization:

- If there is (1) a threat, which is (2) uncertain, then (3) some kind of action (4) is mandatory (Sandin 1999, p. 891).

Accordingly, there are four basic dimensions of the principle:

1. The *threat dimension* that refers to one or other undesired possible state of the world
2. The *uncertainty dimension* that expresses our (lack of) knowledge of these possible states of the world
3. The *action dimension* that concerns what response to the threat is prescribed
4. The *command dimension* that states what the status of the action is (Sandin 1999, pp. 890–895).

The content of these dimensions vary considerably between different versions of the precautionary principle. But arguably, many disputes concerning the exact meaning of the principle could be reduced to disagreements on the proper range of the variables (1)–(4) in this structural schema.

A slight modification of the above is to think different formulations of the precautionary principle as a function of a trigger condition and precautionary response. When a situation fulfills the prerequisites described by the trigger condition, the stated precautionary response should be taken (or taking the precautionary response is justified). The trigger is twofold. It consists of *damage* and *knowledge thresholds* which determine the necessary and jointly sufficient preconditions for the application of precaution. A damage threshold specifies the relevant harmful or otherwise undesirable outcomes. They typically include serious and/or irreversible environmental damage and health hazards. A knowledge threshold defines the required level of scientific understanding of an identified threat. According to a common view based on a decision-theoretic classification, the principle can be applied when the (objective) probability of a threat cannot be established or when its magnitude or severity is uncertain or contested. If the probability of a threat is known and/or relatively high, it is common to talk about taking preventions rather than precautions. (The distinction between the prevention principle and the precautionary principle is more complicated in legal practice.) A broader view, which rests on the level of scientific understanding, states that taking precautionary measures is well-founded when a threat is poorly understood in scientific terms, or when there are scientific discrepancies and/or disagreements on the risk in question.

Precautionary response means taking preemptive measures. These may take the form of outright bans or phaseouts, moratoria, premarket testing, labeling, and requests for extra scientific information before proceeding. Another kind of precautionary response might be establishing new precautionary risk assessment methodologies. The focus is then not only on how to deal with the identified threats, but also on the methods to anticipate and assess threats in the first place. (When these methodologies are in use, they may be considered to belong to the trigger side as they change the trigger condition for taking precautions. As an example of precautionary risk assessment methodologies see Tickner 2003. For an analysis of the narrow and broad precautionary policies implemented within the EU see Levidow et al. 2005. For discussion on different kinds of precautionary measures see, e.g., Whiteside 2006, pp. 52–55).

## The Weak/Strong Distinction

---

It has become common to distinguish between weak and strong interpretations of the precautionary principle. Scholars have employed the distinction in order to evaluate different understandings/formulations of the precautionary principle (e.g., Harris and Holm 2002; Hughes 2006; Morris 2000; Soule 2002; Sunstein 2005, Chap. 1). It is frequently referred to in academic literature, environmental policy reports, and the public discussion on the principle.

Morris defines the strong form as follows: “take no action unless you are certain that it will do no harm.” According to its weak counterpart, lack of certainty is not a justification for preventing an action that might be harmful (Morris 2000, p. 1). Business ethicist Edward Soule argues that the weak form provides regulators with the authority to override other factors and make environmental risk the deciding concern. This is however optional because it does not obligate regulators to treat environmental risk in this way. The strong interpretation, in its turn, restricts regulators to consider environmental risk in isolation from possible benefits. Taking precautions is not optional but mandatory (Soule 2002, pp. 18–19, 22, 24).

While the Rio formulation (UNCED 1992) is typically thought to represent a paradigm example of the weak form of the precautionary principle, the *Wingspread Statement* (1998) is the most frequently provided example of the strong form. Other examples which are commonly subsumed under the weak form include formulations of the precautionary principle found in the *Ministerial Declaration on Sustainable Development in the ECE Region* (1990), the *Communication on the Precautionary Principle* (CEC 2000, p. 4), the *Cartagena Protocol on Biosafety to the Convention on Biological Diversity* (CPB 2000, Article 10), and the *Second North Sea Declaration* (1987, Preamble, Paragraph VII). In contrast, the *Third North Sea Declaration* (1990, Preamble) and the *World Charter for Nature* (1982, Principle 11b) are frequently considered to represent the strong form.

Some authors find both interpretations to be well-grounded. According to Soule (2002, pp. 18, 30), “in domestic contexts the weak formulations are unobjectionable (...) [t]he strong PP [precautionary principle] might seem to require some repairs and not rejection.” Others claim that neither the weak nor the strong form is credible (e.g., Harris and Holm 2002; Morris 2000). Yet, most often, the weak form is argued to be a valid tool of environmental and health risk governance, and it is contrasted with the strong one, which is considered unacceptably extreme, incoherent, or otherwise implausible (e.g., Hughes 2006; Sunstein 2005).

Interestingly, the weak/strong distinction has been employed (or “defined”) in several ways. It is made on the basis of different, sometimes several, criteria instead of one and the same or generally agreed ones. First, the weak/strong distinction is often associated with placement of the burden of proof (e.g., Wiener and Rogers 2002; see also Hohmann 1994). Accordingly, the strong interpretation (S1) says that the proponent of an activity should demonstrate that the activity in question does not lead to an environmental disaster. On the other hand, the weak interpretation (W1) reduces the evidence threshold of a government to interfere in actions of the scientific community, industry, etc. The precautionary principle is a policy instrument which is used to justify restrictions when policymakers have no scientific proof that the action(s) in question would cause harm (Manson 1999).

At other times, the decisive criterion which distinguishes the strong interpretation from the weak one is taken to be the normative status of the prescribed precautionary measures (e.g., Conko 2003; Cameron and Wade-Gery 1995). According to the strong interpretation (S2), there is an obligation to take precautionary measures, whereas the weak counterpart (W2) states merely that precautionary measures are justified (Godard 1997, p. 25). Sometimes the strong form and weak form are also distinguished by referring to the status of cost-benefit analysis (e.g., Myhr and Traavik 2003; see also Soule 2002). The strong interpretation (S3) then implies that non-environmental consequences of taking precautions – e.g., possible economic losses – should not be taken into consideration or that they are always overridden by environmental and health considerations. The prohibition is categorical (Nollkaemper 1996). In contrast, the weak interpretation (W3) requires that the chosen precautionary measures are cost-effective. The known and predicted costs and benefits of different measures and the option of having no measures should be considered (see, e.g., CEC 2000; *Wingspread Statement* 1998). Instead of straightforward bans and moratoria, taking precautionary measures can also mean increased monitoring of an activity, etc.

It has also been thought that the criterion which distinguishes the strong form from the weak one is the status of scientific evidence. The weak interpretation (W4) says that any implementation of precautionary measures presupposes scientific evidence for a hazard which has been identified and assessed in the preceding risk assessment (see Foster et al. 2000; Morris 2000). The strong counterpart (S4) states that scientific proof is not a necessary condition for the application of precautionary measures. Lastly, worth noticing is that some scholars do not employ the strong form and weak form as sharply distinguished groups but as a continuum (e.g., Hughes 2006; Sunstein 2005; Tinker 1996).

## Types of Precautionary Principles

---

Another way of classifying different versions of the precautionary principle is according to what type of rule they express (Sandin 2007). According to this approach, versions of the precautionary principle can express (1) rules of choice, (2) procedural requirements, and (3) epistemic rules or principles.

In some cases, the precautionary principle is formulated as a *rule of choice*. The outcome of such a rule is an action or set of actions that should be chosen, or not chosen (cf. Peterson 2003). The *Wingspread Statement* (1998) version of the precautionary principle is an example of this. It states that some courses of action are not permissible. Those commentators who identify the precautionary principle with the maximin decision rule (Hansson 1997) also see

the principle as a rule of choice. Maximin is a proposed decision rule for decisions under ignorance stating that you should choose the action where the worst outcome is least bad, i.e., maximize the minimum.

In other cases, the precautionary principle is a *procedural requirement*. They do not prescribe what action should be chosen but state conditions for how such a choice should be made. The most widely cited version of the precautionary principle of this kind is the Rio formulation stating that under conditions of scientific uncertainty, “lack of full scientific certainty *shall not be used as a reason* for postponing cost-effective measures to prevent environmental degradation” (UNCED 1992, emphasis added). In this category, we also find burden-of-proof requirements, stating that proponents of a potentially risky activity should bear the burden of proof: This means that those proponents should be required to show that the activity is safe, rather than that the authorities should be required to show that the activity is proven to be hazardous before prohibiting it. Incidentally, such a requirement is also found in the *Wingspread Statement* (1998) which sets out criteria for application of the principle in the paragraphs following the definition.

In the last type of cases, the precautionary principle is an *epistemic rule or principle*. These versions are not about how we should act or how decisions should be made, but about what we should believe. In risk management terminology, this would in practice mean moving precaution from the risk management to risk assessment. Several commentators have been critical of the precautionary principle as an epistemic rule (Harris and Holm 2002; Peterson 2007; Sandin 2007). The main objection is that such a principle would lead us to incorporate a number of false beliefs in our belief system, which would undermine our epistemic basis for decision-making, leading us to bad decisions.

## Arguments Against the Precautionary Principle

---

Several criticisms have been leveled at the precautionary principle. The *argument from vagueness* is one of the most common. It says that the precautionary principle is ill-defined and thus too vacuous to offer any useful guidance for decision-making. Consequently, the principle should be abandoned. Bodansky (1991, p. 5), for example, has argued that the precautionary principle cannot serve as a regulatory standard because it does not specify how much (pre)caution should be taken. Yet he concludes that the principle may play a role in environmental policy as a general goal and stresses the use of discretion in its implementation (Bodansky 1991, p. 43). A more pessimistic conclusion is drawn by Turner and Hartzell (2004, pp. 449, 451, 459) when they claim that “the precautionary principle, in all of its forms, is fraught with vagueness and ambiguity,” that “there is no way of gaining precision and conceptual clarity without sacrificing plausibility,” and that the precautionary principle can serve us neither as a moral principle nor as a decision-making principle (see also Sunstein 2005, pp. 54–55).

The *argument from incoherence* has the following basic logic: incoherent principles should not be used as a basis for societal risk decision-making; the precautionary principle is incoherent; thus, it should be abandoned. Philosopher Gary Comstock (2000) has argued that “[t]he precautionary principle commits us to each of the following propositions: (1) We must not develop GM crops. (2) We must develop GM crops.” In their paper “Extending Human Lifespan and the Precautionary Paradox,” Harris and Holm (2002, see also 1999) similarly claim that the precautionary principle is incoherent and consequently does not provide

the kind of justification (for a precautionary “pause” from proceeding with new technologies) that it is often presumed to offer. Their main argument is that the principle cannot coherently be employed as a decision rule, an epistemic rule, or a moral principle (cf. Sandin 2006). The argument from incoherence is also found in Sunstein (2005) and Peterson (2006).

It has also been argued that the implementation of the precautionary principle would lead to serious and commonly unwanted consequences, and thus that the principle should be abandoned as a policymaking tool. The *argument from adverse effects* says that, instead of decreasing it, the precautionary principle increases our risk imposition in total. This argument takes several forms. The use of the principle may result in different kinds of adverse effects – directly or indirectly. Precautionary measures taken may in themselves impose a new environmental threat or a health hazard. In his book entitled *The Precautionary Principle: A Critical Appraisal of Environment Risk Assessment*, engineer and policy analyst Indur Goklany (2001, see also 2000) provides us with a detailed analysis as to why the application of the precautionary principle to various contentious environmental issues may result in undesirable effects and increased risk taking. A similar kind of an argument is put forward by policy analysts Henry I. Miller and Gregory Conko who argue that “[i]f the precautionary principle had been applied decades ago to innovations like polio vaccines, and antibiotics, regulators might have prevented occasionally serious, and sometimes fatal, side effects by delaying or denying approval of those products, but that precaution would have come at the expense of millions of lives lost to infectious diseases” (Miller and Conko 2000, p. 100).

Lastly, to argue that the precautionary principle is a value judgment or an “ideology,” that it is unscientific, or that it blurs the boundary between science and policy in an unacceptable way is not uncommon (e.g., Morris 2000; Wildavsky 1996; Gray and Bewers 1996) (For analyses of general arguments against the precautionary principle, see Sandin et al. 2002; Ahteesuu 2007).

## Further Research

---

The precautionary principle has recently received much attention in academic discourse, and this body of literature has virtually exploded over the last decade. Yet there remain areas of research which deserve more thorough scrutiny. First, there are methods of inquiry which have hitherto been insufficiently utilized. In particular, this applies to formal methods. Second, certain topics deserve further study. These include the normative underpinnings of the principle, the status of the principle in scientific risk analysis, and the principle’s relationship with stakeholder/public engagement.

## Formal Methods

---

Comparatively, few works approach the precautionary principle using formal methods (see, e.g., Ready and Bishop 1991; Perrings 1991; Peterson 2006). Some of economics and decision theory’s developments come close to its basic idea. Economist Frank Knight (1921, p. 11, Part III Chaps. VII & VIII) distinguished between quantifiable risks (i.e., uncertainties which can be represented as statistical odds or probabilities) and “true” uncertainties which cannot be quantified (i.e., assessed in this way). The precautionary principle is typically considered to be applicable to situations in which the probability of a risk cannot be assigned. Irrevocability

as a qualification of the damage threshold of the precautionary principle has earlier counterparts. Irreversible consequences in terms of restricting tomorrow's set of opportunities (or decision possibilities) with today's choices have been a subject of close scrutiny and debate in decision-theoretic literature since the early seventies. Similarities can also be found between the precautionary principle and the maximin decision rule, and some early commentaries treat them as amounting to the same thing (Hansson 1997). Others have interpreted the precautionary principle as minimax regret (Chisholm and Clarke 1993). This is certainly an area in which considerable work remains to be done.

## Normative Underpinnings

---

As a principle of practical decision-making, the precautionary principle may be justified on the basis of ethical and sociopolitical grounds and/or as a form of rational action. The application of the principle is fundamentally a normative (and political) choice. The degree to which we are prepared to take precautions is related to the values which we attach to the nature and human well-being. Notwithstanding, normative underpinnings of the precautionary principle have received only little attention. Although the importance of ethical discussion has been underlined in several occasions, there are only a few published papers on the issue. Susan Carr (2002) at the Open University in the UK has criticized the EU's negligence of the value-based aspects of the principle. The Commission of the European Communities has emphasized the scientific aspects of precautionary decision-making and ignored almost totally the justification of its basic values. Environmental scientist and ethicist Marc A. Saner (2002) relates the precautionary principle to the main approaches in the Western ethical traditions, in particular to those of virtue ethics, deontology, and utilitarianism (On normative underpinnings, see also Jensen 2002; Munthe 2011; Parker 1998; von Schomberg 2006). In policymaking, the precautionary principle has been invoked to justify a wide range of policies – sometimes even mutually contradictory ones (see Levidow et al. 2005). This has often been done without an explicitly stated normative framework.

## Risk Analysis

---

The precautionary principle has typically been thought to present a risk management principle or tool (e.g., CEC 2000, pp. 3, 13). This is reflected, for example, by the following statements: "Precautionary principles have been proposed as a fundamental element of sound risk management" (Löfstedt et al. 2002, p. 381), and the principle "is invoked in the process of risk management" (Rogers 2001, p. 1). In practice, this means that the precautionary principle can be applied when a risk has been identified in the preceding risk assessment, but a considerable amount of uncertainty remains (i.e., the probability of the risk cannot be quantified and/or the magnitude of the risk is unknown). Yet other standpoints have also been suggested. Some authors argue that the precautionary principle should already be taken into consideration at the level of risk assessment. According to this position, the principle works not only as a risk management principle/tool, but it also affects the way in which risk assessment is conducted (Goldstein and Carruth 2004, pp. 491–493; see also Levidow et al. 2005, pp. 268–269). New notions such as "precautionary appraisal" (Klinke et al. 2009) and "precautionary assessment" (Tickner 2003) have been introduced. The relevance of the precautionary principle to risk assessment deserves more study.

## Participatory Decision-Making Practices

Participatory decision-making is frequently connected with the precautionary principle. In their report *Late Lessons from Early Warnings*, European Environment Agency (EEA 2001, p. 17) argues that implementing the principle calls for interest groups participation which “should be at an early stage, broadly drawn, and carried down to the appropriate local level” (See also Fisher and Harding 1999, p. 290; Funtowicz and Ravetz 1993). Yet, what is the exact nature of this link between precaution and stakeholder/public engagement remains unclear. The question in itself has been somewhat neglected in the related theoretical literature. Biologist John Lemons, philosopher Kristin Shrader-Frechette, and philosopher Carl Cranor suggest that a “fundamental dilemma surrounding (...) the use of a precautionary approach is how to balance the need for expert scientific knowledge with the need to involve the public in the decision-making process” (Lemons et al. 1997, pp. 233–234). According to environmental law scholar Elizabeth Fisher (2001, p. 320), “while there is little agreement over what the nature of that participation should be it is clearly an important part of a precautionary decision-making process.” Even though several comments and some proposals for inclusive precautionary risk governance frameworks have been put forward, relatively few analyses of the relationship between precaution and engagement is available (see, e.g., Fisher and Harding 1999; Raffensberger and Barrett 2001; see also Ravetz 2004; for an analysis of precaution and democratic deliberation see Whiteside 2006, Chap. 5).

## References

- Adams MD (2002) The precautionary principle and the rhetoric behind it. *J Risk Res* 5:301–316
- Ahteeensuu M (2007) Defending the precautionary principle against three criticisms. *Trames J Humanit Soc Sci* 11(4):366–381
- Bodansky D (1991) Scientific uncertainty and the precautionary principle. *Environment* 33(7):4–5, 43–44
- Boehmer-Christiansen S (1994) The precautionary principle in Germany: enabling government. In: O’Riordan T, Cameron J (eds) *Interpreting the precautionary principle*. Cameron May, London
- Cameron J, Wade-Gery W (1995) Addressing uncertainty: law, policy and the development of the precautionary principle. In: Dente B (ed) *Environmental policy in search of new instruments*. Kluwer, Dordrecht
- Carr S (2002) Ethical and value-based aspects of the European Commission’s precautionary principle. *J Agric Environ Ethics* 15:31–38
- CEC (Commission of the European Communities) (2000) Communication from the Commission on the precautionary principle (Brussels, 2 Feb 2000 COM[2000]1)
- Chisholm A, Clarke HR (1993) Natural resource management and the precautionary principle. In: Dommen E (ed) *Fair principles for sustainable development: essays on environmental policy and developing countries*. Edward Elgar, Aldershot, pp 109–122
- Comstock G (2000) Are the policy implications of the precautionary principle coherent? AgBioWorld 2000. [http://www.agbioworld.org/newsletter\\_wm/index.php?cascid=archive&newsid=134](http://www.agbioworld.org/newsletter_wm/index.php?cascid=archive&newsid=134). Accessed 10 May 2011
- Conko G (2003) Safety, risk and the precautionary principle: rethinking precautionary approaches to the regulation of transgenic plants. *Transgenic Res* 12:639–647
- CPB (Secretariat of the Convention on Biological Diversity) (2000) Cartagena protocol on biosafety to the convention on biological diversity: text and annexes, Montreal
- Cranor CF (2003) What could precautionary science be? Research for early warnings and a better future. In: Tickner J (ed) *Precaution, environmental science, and preventive public policy*. Island Press, Washington, DC
- de Sadeleer N (2002) Environmental principles: from political slogans to legal rules. Oxford University Press, New York
- de Sadeleer N (ed) (2006) *Implementing the precautionary principle: approaches from the Nordic Countries, EU and USA*. Earthscan, London

- EEA (European Environment Agency) (2001) Late lessons from early warnings: the precautionary principle 1896–2000. [http://reports.eea.eu.int/environmental\\_issue\\_report\\_2001\\_22/en/Issue\\_Report\\_No\\_22.pdf](http://reports.eea.eu.int/environmental_issue_report_2001_22/en/Issue_Report_No_22.pdf). Accessed 10 May 2011
- Fisher E (2001) Is the precautionary principle justiciable? *J Environ Law* 13(3):315–334
- Fisher E, Harding R (1999) The precautionary principle: towards a deliberative, transdisciplinary problem-solving process. In: Harding R, Fisher E (eds) *Perspectives on the precautionary principle*. Federation, Sydney
- Fitzmaurice M (2009) Contemporary issues in international environmental law. Edward Elgar, Cheltenham
- Foster K, Vecchia P, Repacholi MH (2000) Science and the precautionary principle. *Science* 288:979–981
- Freestone D, Hey E (eds) (1996) The precautionary principle and international law: the challenge of implementation. Kluwer Law International, The Hague
- Funtowicz SO, Ravetz JR (1993) Science for the post-normal age. *Futures* 25(7):739–755
- Gardiner SM (2006) A core precautionary principle. *J Polit Philos* 14(1):33–60
- Godard O (1997) Introduction générale. In: Godard O (ed) *Le principe de précaution dans la conduite des affaires humaines*. MSH et INRA, Paris
- Goklany IM (2000) Applying the precautionary principle in a broader context. In: Morris J (ed) *Rethinking risk and the precautionary principle*. Butterworth-Heinemann, Oxford, pp 189–228
- Goklany IM (2001) The precautionary principle: a critical appraisal of environment risk assessment. Cato Institute, Washington, DC
- Goldstein B, Carruth RS (2004) The precautionary principle and/or risk assessment in world trade organization decisions: a possible role for risk perception. *Risk Anal* 24(2):491–499
- Goodin RE (1980) No moral nukes. *Ethics* 90:417–449
- Graham JD, Hsia S (2002) Europe's precautionary principle: promise and pitfalls. *J Risk Res* 5(4):371–390
- Gray JS, Bewers M (1996) Towards a scientific definition of the precautionary principle. *Mar Pollut Bull* 32(11):768–771
- Hansson SO (1997) The limits of precaution. *Found Sci* 2(2):293–306
- Harris J, Holm S (2002) Extending human lifespan and the precautionary paradox. *J Med Philos* 27(3):355–368
- Hohmann H (1994) Precautionary legal duties and principles of modern international environmental law: the precautionary principle: international environmental law between exploitation and protection. *International Environmental Law and Policy Series*, London
- Holm S, Harris J (1999) Precautionary principle stifles discovery. *Nature* 400:398
- Hughes J (2006) How not to criticize the precautionary principle? *J Med Philos* 31:447–464
- Jensen K (2002) The moral foundation of the precautionary principle. *J Agric Environ Ethics* 15:39–55
- Jonas H (1984) *The imperative of responsibility: in search of an ethics for the technological age*. University of Chicago Press, Chicago
- Jordan A, O'Riordan T (1999) The precautionary principle in contemporary environmental policy and politics. In: Raffensberger C, Tickner J (eds) *Protecting public health and the environment: implementing the precautionary principle*. Island Press, Washington, DC
- Karlsson M (2006) The precautionary principle, Swedish chemicals policy and sustainable development. *J Risk Res* 9(4):337–360
- Klinke A, Renn O, Dreyer M, Losert C (2009) Project outline and conceptual considerations for a general model of precautionary risk regulation. In: Renn O, Schweizer P-J, Müller-Herold U, Stirling A (eds) *Precautionary risk appraisal and management: an orientation for meeting the precautionary principle in the European Union*. Europäische Hochschulverlag, Bremen
- Knight FH (1921) *Risk, uncertainty and profit*. Hart, Schaffner & Marx/Houghton-Mifflin Co., Boston. The Riverside Press, Cambridge (Reprinted University of Chicago Press 1971)
- Lemons J, Shrader-Frechette K, Cranor C (1997) The precautionary principle: scientific uncertainty and type I and type II errors. *Found Sci* 2:207–236
- Levidow L, Carr S, Wield D (2005) European Union regulation of agri-biotechnology: precautionary links between science, expertise and policy. *Sci Public Policy* 32(4):261–276
- Löfstedt RE, Fischhoff B, Fischhoff IR (2002) Precautionary principles: general definitions and specific applications to genetically modified organisms. *J Policy Anal Manage* 21(3):381–407
- Mace PM, Sissenwine MP (2002) Coping with uncertainty: evolution of the relationship between science and management. In: Berkson JM, Kline LL, Orth DJ (eds) *Incorporating uncertainty into fishery models*. American Fisheries Society, Bethesda, pp 9–28
- Manson NA (1999) The precautionary principle, the catastrophe argument, and Pascal's wager. *J Ends Means* 4:12–16
- Manson NA (2002) Formulating the precautionary principle. *Environ Ethics* 24:263–274
- Martin PH (1997) If you don't know how to fix it, please stop breaking it! *Found Sci* 2:263–292
- Mascher S (1997) Taking a 'precautionary approach': fisheries management in New Zealand. *Environ Plan Law J* 14:70–79
- Miller HI, Conko G (2000) Genetically modified fear and the international regulation of biotechnology. In:

- Morris J (ed) Rethinking risk and the precautionary principle. Butterworth-Heinemann, Oxford, pp 84–104
- Ministerial Declaration of the Second International Conference on the Protection of the North Sea (London, 25 Nov 1987)
- Ministerial Declaration of the Third International Conference on the Protection of the North Sea (The Hague, 8 Mar 1990)
- Ministerial Declaration on Sustainable Development in the ECE Region (Bergen, 16 May 1990)
- Morris J (2000) Defining the precautionary principle. In: Morris J (ed) Rethinking risk and the precautionary principle. Butterworth-Heinemann, Oxford, pp 1–21
- Munthe C (2011) The price of precaution and the ethics of risk. Springer, Berlin
- Myers NJ (2006) Introduction. In: Myers NJ, Raffensperger C (eds) Precautionary tools for reshaping environmental policy. MIT Press, Cambridge
- Myhr AI, Traavik T (2003) Genetically modified (GM) crops: precautionary science and conflicts of interests. *J Agric Environ Ethics* 16:227–247
- Nollkaemper A (1996) ‘What you risk reveals what you value’ and other dilemmas encountered in the legal assaults on risks. In: Freestone D, Hey E (eds) The precautionary principle and international law: the challenge of implementation. Kluwer Law International, The Hague, pp 73–94
- O’Riordan T, Cameron J, Jordan A (eds) (2002) Reinterpreting the precautionary principle. Cameron May, London
- Ortega Vicuña F (1999) The changing international law of high seas fisheries. Cambridge University Press, Cambridge
- Ozonoff D (1999) The precautionary principle as a screening device. In: Raffensperger C, Tickner J (eds) Protecting public health and the environment: implementing the precautionary principle. Island Press, Washington, DC
- Parker J (1998) Precautionary principle. In: Chadwick R (ed) Encyclopedia of applied ethics, vol 3. Academic Press, San Diego
- Pearce DW et al (1980) The preconditions for achieving consensus in the context of technological risk. In: Dierkes M, Edwards S, Coppock R (eds) Technological risk: its perception and handling in the European Community. Gunn & Hain, Cambridge
- Perrings C (1991) Reserved rationality and the precautionary principle. In: Costanza R (ed) Ecological economics: the science and management of sustainability. Cambridge University Press, New York, pp 153–166
- Peterson M (2003) Transformative decision rules. *Erkenntnis* 58:71–85
- Peterson M (2006) The precautionary principle is incoherent. *Risk Anal* 26(3):595–601
- Peterson M (2007) Should the precautionary principle guide our actions or our beliefs? *J Med Ethics* 33(1):5–10
- Raffensperger C, Barrett K (2001) In defense of the precautionary principle. *Nature Biotechnol* 19:811–812
- Randall A (2011) Risk and precaution. Cambridge University Press, Cambridge
- Ravetz J (2004) The post-normal science of precaution. *Futures* 36:347–357
- Ready RC, Bishop RC (1991) Endangered species and the safe minimum standard of conservation. *Am J Agric Econ* 72(2):309–312
- Resnik DB (2003) Is the precautionary principle unscientific? *Stud Hist Philos Biol Biomed Sci* 34:329–344
- Rogers MD (2001) Scientific and technological uncertainty, the precautionary principle, scenarios and risk management. *J Risk Res* 4(1):1–15
- Sandin P (1999) Dimensions of the precautionary principle. *Hum Ecol Risk Assess* 5:889–907
- Sandin P (2004a) Better safe than sorry: applying philosophical methods to the debate on risk and the precautionary principle. Theses in philosophy from the Royal Institute of Technology (Academic dissertation), Stockholm
- Sandin P (2004b) The precautionary principle and the concept of precaution. *Environ Values* 13:461–475
- Sandin P (2006) A paradox out of context: Harris and Holm on the precautionary principle. *Camb Q Healthc Ethics* 15(2):175–183
- Sandin P (2007) Common-sense precaution and varieties of the precautionary principle. In: Lewens T (ed) Risk: philosophical perspectives. Routledge, London, pp 99–112
- Sandin P, Peterson M, Hansson SO, Rudén C, Juthe A (2002) Five charges against the precautionary principle. *J Risk Res* 5:287–299
- Saner M (2002) An ethical analysis of the precautionary principle. *Int J Biotechnol* 4(1):81–95
- Soule E (2002) Assessing the precautionary principle in the regulation of genetically modified organisms. *Int J Biotechnol* 4(1):18–33
- Starr C (2003) The precautionary principle versus risk analysis. *Risk Anal* 23(1):1–3
- Sunstein C (2005) Laws of fear: beyond the precautionary principle. Cambridge University Press, Cambridge
- Tickner J (2003) Precautionary assessment: a framework for integrating science, uncertainty, and preventative policy. In: Tickner J (ed) Precaution, environmental science, and preventive public policy. Island Press, Washington, DC
- Tinker C (1996) State responsibility and the precautionary principle. In: Freestone D, Hey E (eds) The

- precautionary principle and international law: the challenge of implementation. Kluwer Law International, The Hague
- Trouwborst A (2002) Evolution and status of the precautionary principle in international law. Kluwer Law International, London
- Turner D, Hartzell L (2004) The lack of clarity in the precautionary principle. *Environ Values* 13:449–460
- UNCED Rio Declaration on Environment and Development (United Nations Conference on Environment and Development, Rio de Janeiro, 3–14 June 1992)
- United Nations Environment Programme (1982) World charter for nature: United Nations General Assembly (Resolution 37/7). UNEP, Nairobi, 28 Oct 1982
- VanderZwaag D (2002) The precautionary principle and marine environmental protection: slippery shores, rough seas, and rising normative tides. *Ocean Dev Int Law* 33:165–188
- von Schomberg R (2006) The precautionary principle and its normative challenges. In: Fisher E, Jones J, von Schomberg R (eds) *Implementing the precautionary principle: perspectives and prospects*. Edward Elgar, Cheltenham, pp 19–41
- Whiteside KH (2006) *Precautionary politics: principle and practice in confronting environmental risk*. MIT Press, Cambridge
- Wiener JB, Rogers MD (2002) Comparing precaution in the United States and Europe. *J Risk Res* 5:317–349
- Wildavsky A (1996) *But is it true? A citizen's guide to environmental health and safety issues*. Harvard University Press, Cambridge
- Wingspread Statement on the Precautionary Principle (1998), Racine, WI, 26 Jan 1998

# 39 The Capability Approach in Risk Analysis

Colleen Murphy · Paolo Gardoni

Texas A&M University, College Station, TX, USA

<b>Introduction</b> .....	<b>980</b>
<b>History</b> .....	<b>981</b>
Defining Capability .....	981
The Capability Approach and Gaging Capabilities .....	983
<b>Current Research</b> .....	<b>984</b>
Use of the Capability Approach in Risk Theory .....	984
Risk Determination .....	984
Risk Evaluation .....	989
Risk Management .....	991
Use of Risk Theory in the Capability Approach .....	993
<b>Further Research</b> .....	<b>994</b>
<b>Conclusion</b> .....	<b>995</b>

**Abstract:** The standard meaning of risk adopted by risk analysts from a broad range of fields is that risk is the probability that a certain set of consequences will occur given a hazardous scenario. Risk analysis is the process of determining the probability of occurrence and consequences as well as evaluating the determined risks. Capability refers to the genuine opportunity that an individual has to do and become things of value, such as being educated and maintaining bodily integrity. Such doings and beings are called functionings. A capability approach provides a distinct evaluative space for conceptualizing and judging states of affairs on the basis of how the capabilities of individuals are affected. This chapter examines the reciprocal contributions of a capability approach and risk theory. A capability approach has been used to enrich risk theory in three ways. First, it has been argued that the consequences component of risk should be conceptualized and assessed in terms of the impact of a hazardous scenario on capabilities, instead of either resources or utility. Second, instead of evaluating risks on the basis of either public or expert judgment, risks should be judged acceptable or tolerable on the basis of whether certain threshold levels of capabilities are maintained. Third, a capability approach provides a distinct alternative to a prominent way of managing risks, cost-benefit analysis. Instead of comparing alternative policies on the basis of their relative advantages and disadvantages, policies should be evaluated on the basis of whether they address unacceptable or intolerable risks and on the basis of their likely affectability. Likely affectability considers the dollar per unit change in the expected impact of a hazardous scenario on capabilities. Risk theory has enriched the capability approach in two ways. First, considering risk has highlighted an important dimension of capability, security, not previously recognized. Second, attempts to operationalize a capability approach to risk have led to the development of a novel way to assess capabilities, and not simply actual functioning achievements.

## Introduction

---

In this chapter we discuss a capability approach in risk analysis and discuss how risk analysis has enriched theorizing about capabilities. The standard meaning of risk adopted by risk analysts from a broad range of fields is that risk is the probability that a certain set of consequences will occur given a hazardous scenario (Kaplan and Garrick 1981; Hansson 2007b). This definition is more specific than the broader definition of risk found in the philosophical literature according to which risk is “a possible scenario in which adverse events take place” (Hansson 2007b). Risk analysis is the process of determining the probability of occurrence and consequences as well as evaluating the specified risks. There are three different components of risk analysis: risk determination, risk evaluation, and risk management. Capability refers to the genuine opportunity that an individual has to do and become things of value, such as being educated and maintaining bodily integrity. The capability approach provides a distinct evaluative space for conceptualizing and judging states of affairs on the basis of how the capabilities of individuals are affected.

There are four sections in this chapter. The first defines capability and a capability approach and discusses how capabilities are assessed. The second provides an overview of how the capability approach has been applied to risk analysis as well as how ideas from risk analysis have been used to deepen our understanding of capabilities. The third outlines two main areas for further research in a capability approach to risk.

## History

### Defining Capability

There are many valuable states and activities, such as being adequately nourished, being educated, and participating in the political life of a community, which a given individual may or may not be in a position to achieve. In the capability approach pioneered by Amartya Sen (1989, 1992, 1993, 1999a, b) and Martha Nussbaum (2000a, b, 2001) such states and activities are called functionings. An individual enjoys a particular capability if he or she has a genuine opportunity to achieve a specific valuable functioning. In practice, capabilities are interdependent. An individual might have a genuine opportunity to have a rewarding career and a large family, but not have an opportunity to achieve both at the same time. An individual's general capability captures this interconnectedness. An individual's general capability is a function of the various combinations or vectors of functionings that he or she has a genuine opportunity to achieve (Sen 1992).

Two general factors influence the capability of an individual: what he or she has and what he or she can do with what he or she has (Wolff and de-Shalit 2007). The resources of an individual provide information about what he or she has. Internal resources refer to the assets an individual possesses, including skills, talents, and psychological well-being. External resources encompass the other means available to him or her, including income, wealth, and the support of family. For example, whether an individual has a genuine opportunity to be educated depends on her mental resources and talents. However, it may also depend on income, insofar as there is tuition or costs for lodging, meals, and books associated with schooling; moreover, schooling often reduces an individual's ability to earn income in the short term. Family support for education is also critical; families must be willing, for example, to allow a child to attend school instead of working inside or outside of the home. The resources of an individual are not the only factor that influences what genuine opportunities are available to him or her. There are variations in what individuals can achieve with a given set of resources (Sen 1993, 1999b). Equally salient for determining the genuine opportunities of an individual is the social and material structure within which he or she acts. Customs and traditions, laws, the physical infrastructure of a community, and language are all examples of some of the dimensions of this social and material structure. It is this structure that influences what an individual can do or become with the resources at his or her disposal. An individual may have a set of talents and skills. However, whether an individual can become educated given his or her talents and skills depends on the condition of the physical infrastructure (e.g., whether there are roads or paths through which he or she can access a school), social norms, and laws (e.g., whether there are formal or informal restrictions on him or her attending school). Such factors impact whether and how resources contribute to achieved functionings (Robeyns 2005, p. 99).

An individual's capability is morally significant because of its connection with individual well-being. According to Sen (2009), well-being can refer to either the "promotion of an individual's well-being" or the promotion of an individual's agency goals. Agency goals are a broader category of goals, encompassing everything an individual has reason to pursue. Some of the goals of an individual may include his or her own well-being, and so an increase in well-being will be an increase in agency achievement. However, some agency goals may not be connected to an individual's well-being; indeed the achievement of some agency goals can

be at the expense of individual well-being. Using the freedom/achievement distinction, an individual's well-being might relate to well-being achievement or well-being freedom. Keeping these distinctions in mind, Sen claims that capability, broadly understood as the dimension of freedom concerned with substantive opportunities, can include both well-being freedom and agency freedom. He argues that more capability often enhances well-being, when well-being is defined in terms of agency. However, because that same agency freedom can be in tension with, or indeed go against, individual well-being freedom depending on the goals a particular agent has, capability may not always enhance well-being understood as well-being freedom. Capabilities provide information about the opportunity individuals have (or lack) to exercise their agency in shaping their lives.

Conceptualizing and assessing individual well-being in terms of capabilities differs importantly from two common metrics for assessing well-being: resources and utility. Consider resources, or evaluating well-being on the basis of the market commodities (e.g., income, resources) or primary goods (e.g., wealth, income, social bases of self-respect) an individual enjoys. Resources are one necessary condition for the well-being of individuals. As we saw earlier, resources influence the capabilities of individuals. An increase in resources can also be the consequence of certain functioning achievements. However, one problem with the resource framework is that it mistakes means for ends. Resources are taken to be intrinsically good, instead of instruments for the achievement of what is intrinsically good, namely, capabilities. Typically, we care about income not for its own sake but because of what it allows us to do and become. A second problem with the resource-based framework is that it ignores interpersonal variability, or the different rates at which diverse individuals can use resources to achieve a particular functioning. As David Crocker (2008, p. 114) writes, "Due to variations among individuals, the same commodity either may help some and harm others or may promote the well-being of some a lot and of others only a little. Although food intake normally will enhance human functioning, it will kill the person choking on a fish bone. To function well, Milo the wrestler needs on the one hand, more food than the infant and the disabled and, on the other hand, less food than a wrestler of a similar size but stricken with parasites."

Another method for assessing individual well-being is in terms of utility. Welfare economics is the discipline devoted to the assessment of the goodness of states of affairs and policies in terms of their impact on well-being. Well-being/advantage is usually defined in terms of utility. Utility is defined as happiness, where happiness is often understood as desire fulfillment (Sen 2009). The utility approach avoids one weakness with the resource framework: it does not fetishize goods but makes individuals the fundamental unit of concern (Crocker 2008). However it has important limitations. First, happiness and desire fulfillment are necessary but not sufficient for well-being. Individuals who are persistently deprived may adapt to their circumstances to make life tolerable, learning to "take pleasure in small mercies" and refusing to desire or hope for change in their circumstances. If we assess the well-being/advantage of such individuals on the basis of their happiness alone, then we would fail to get an accurate picture of their actual disadvantage. Conversely, "people may have well-being and even opulence (be 'well off') and yet be unhappy and frustrated; their unfulfilled desires may be for rare Rioja wine and top-of-the-line Mercedes" (Crocker 2008). As these examples illustrate, the informational basis of well-being/advantage in welfare economics is incomplete. It should be broadened to include factors such as substantive opportunities, negative freedoms, and

human rights. Omitting this information prevents us from making important distinctions in our judgments of the relative advantage of individuals who enjoy the same level of happiness, but differ dramatically along these other dimensions. Omitting this information also leads to distorted assessments.

## The Capability Approach and Gaging Capabilities

---

The capability approach offers an evaluative space in terms of which judgments about states of affairs should be made. For example, in a capability approach to development, “[T]he purpose of development is to improve human lives by expanding the range of things that a person can be and do, such as to be healthy and well nourished, to be knowledgeable, and to participate in community life” (Fukuda-Parr and Kumar 2003). Societies are assessed as more or less developed from this perspective, then, not on the basis of their wealth but rather on the basis of the degree to which they ensure the ability of individuals to be healthy, knowledgeable, and a participant in their community. The capability approach is used in a wide variety of fields including social choice theory, welfare economics, development ethics and economics, moral and political theory, disaster studies, and education (Robeyns 2006; Gardoni and Murphy 2010).

To be able to use the approach to evaluate different states of affairs, it must be possible to assess the capabilities of individuals. The earliest attempt to operationalize the capability approach was in development economics. The United Nations and development agencies now assess development on the basis of the functionings achievement of members of a country. The Human Development Report (*HDR*), published annually since 1990, uses the Human Development Index (*HDI*) to assess development. The *HDI* continues to provide a prominent model for assessing functionings achievements in other fields. In the *HDI*, the functionings used for assessing development include the ability to live a long and healthy life, the opportunity to be knowledgeable, and the ability to have a decent standard of living. Capabilities are not directly observable or quantifiable. In practice, indicators are used to try to indirectly gage either actual functionings achievements and/or capability, the freedom to achieve functionings (Raworth and Stewart 2003; United Nations Development Program 2007). Indicators are used to quantify each functioning. To illustrate, one indicator for the ability to live a long and healthy life is life expectancy at birth. For the opportunity to be knowledgeable the adult literacy rate is used. Gross Domestic Product (*GDP*) per capita is an indicator selected for the ability to have a decent standard of living. The data collected for each indicator is compared to a scale of minimum and maximum values (goalposts). In the case of life expectancy, the average life expectancy rate for individuals in a given country is compared to the minimum value of 25 years and the maximum value of 85 years (United Nations Development Program 2000). This comparison, also called normalization or scaling, gives some context and meaning for the specific life expectancy rate of a given country, providing a picture of what the level of functioning achievement for that indicator is (Jahan 2003). The goalposts represent reasonable minimum and maximum values based on an analysis of historical data (Raworth and Stewart 2003). Finally the normalized or scaled values of the individual indicators are combined in an unweighted average. No weights are typically used because each functioning is taken to be equally important.

## Current Research

---

In this section we first consider how the capability approach has been applied to risk analysis. We then discuss two important ways in which the concept of capability has been enriched by risk studies.

### Use of the Capability Approach in Risk Theory

---

The study of risk is highly interdisciplinary. A diverse range of risks are the subject of examination, including those associated to environmental, technological, man-made, and natural hazards. This section describes the contributions that the Capability Approach has made to risk analysis. We consider in particular the contributions to the three components of risk analysis: risk determination, risk evaluation, and risk management.

#### Risk Determination

Risk determination refers to the quantification of the probability of occurrence of a particular hazard and its associated consequences. Our focus in this section is not on the probability component of risk, for which there are a wide variety of methods of quantification available. For a description of such approaches see Paté-Cornell (1996), Haimes (2004), and Cullen and Small (2004). Rather, our interest is in how the consequence components of risk are often characterized and assessed. The first contribution of the capability approach to risk theory is in our understanding of how the consequences from hazardous scenarios should be conceptualized and assessed. Below we first describe two common approaches to conceptualizing and quantifying the consequence components of risk and their limitations. We then discuss how the capability approach has been proposed as an alternative, and how this approach avoids the limitations with existent approaches.

#### Limitations with Common Frameworks to Risk Determination

Risk analysts and those who provide the inputs for a risk analysis, including engineers and social scientists, have historically adopted a broad resource-based approach to the consequence component of risk (Rowe 1980; Vose 2000; Bedford and Cooke 2001; Haimes 2004). In one framework, the numbers of resources of various kinds that are lost, including structures, money, time (in terms of delays in construction), and/or individuals are counted. Counting the consequences of risks in this manner has a number of limitations. First, a concentration on resources lost does not provide information about how the lives of individuals are affected. As we noted in the discussion in the previous section, there are different rates at which a given resource can generate an opportunity to achieve a valuable functioning. Thus, identifying the number of structures damaged does not alone tell us what that loss means for the individuals in the community affected. Losses of a given structure can have a greater or lesser impact depending on whether the structure is a hospital, an empty building, or an apartment complex, for example. Second, defining consequences in terms of resources implicitly treats what are important means to achieving well-being as ends in themselves. Third, risk analysts, engineers, and social scientists increasingly recognize the need to consider consequences beyond those traditionally taken into account in risk analysis. However, the resource framework does not provide a principled way to demarcate what losses should count or are salient for purposes of

risk analysis. Indeed, consequences are often selected in practice on the basis of their ease of quantification (Murphy and Gardoni 2006; Gardoni and Murphy 2009).

A second alternative framework evaluates consequences on the basis of the utility lost in a given hazardous scenario. Utility is commonly determined by examining the preferences or choices of individuals. An underlying assumption is that the satisfaction of preferences increases welfare. Preferences or choices are assessed in monetary terms. Thus, all consequences of interest (e.g., the loss of a human life, harms to the environment) are converted into a monetary figure. This figure is determined on the basis of individual “willingness to pay” (Sunstein 2005). That is, the cost associated with a risk is based on an examination of what individuals would need to be paid to be willing to be exposed to a certain risks. These figures are characteristically based on actual market activity. As Cass Sunstein (2005) describes it, “Suppose that people must be paid \$600, on average, to eliminate risks of 1/10,000; suppose, for example, that workers who face risks of that magnitude generally receive \$600 in additional wages each year. If so, the VSL [Value of a Statistical Life] would be said to be \$6 million.” Contingent valuation surveys are used in cases where there is no market evidence on which to make an assessment. For example, to assess the loss of a coral reef, individuals may be asked what they are willing to pay to salvage coral reefs (Sunstein 2005; Hansson 2007a).

A number of objections have been raised to conceptualizing consequences in terms of utility. Most fundamentally, assigning costs to certain consequences on the basis of market choices and preferences does not accurately convey the costs of certain consequences, either over- or under-estimating their significance. This is so for a number of reasons. First, individuals may have misperceptions about or be indifferent toward risks they face. One common case of misperception of risks is that rare, dramatic risks might be feared more than ordinary risks. In the workplace, individuals may not be fully knowledgeable about the occupational risks they face when accepting a position; when they do learn of certain risks the costs to leave may be grave, including loss of seniority or pension if they seek employment elsewhere (Anderson 1988). Environmental costs may not be a source of grave concern, in part based on a failure to understand the significance of the environment to us. As a result, the costs assigned may lead to a corresponding over- or under-valuation of certain kinds of losses (Slovic 1987; May 2001; Murphy and Gardoni 2006).

Second, even when individuals have an accurate perception of the risks they face and are not indifferent to their consequences, the willingness to pay amount may not capture the way individuals value certain costs, such as the loss of their lives. Risks that are job-related may not be voluntarily accepted, but rather chosen because they were the only form of employment available (Wolff and de-Shalit 2007). As Elizabeth Anderson (1988) writes, “Workers with families commonly do not regard their choices about risks to their lives as acts of consumers out to maximize their personal utilities. Rather, they see their choices as attempts to discharge their responsibilities to their families. Bound by their responsibilities to others, they do not feel free to risk their lives at will for pay, as they sometimes acknowledge having done in their youth. But neither do they feel free to jeopardize their families’ means of livelihood and future prospects by simply quitting their hazardous jobs and risking long-term unemployment.” Thus, choices to accept certain wages may not reflect an individual’s judgment about the value of his/her life, but rather the judgment about how it is necessary to fulfill certain responsibilities.

An additional limitation is that the utilitarian framework does not focus attention on what should be of fundamental interest, namely, how the genuine opportunities of individuals will

change if certain hazards are realized. Instead of valuing and prioritizing the promotion of the agency of individuals, the utilitarian framework treats the losses of human lives as a market commodity. Furthermore, information about the monetary costs of consequences does not translate into information about how the freedom of individuals to do and become things of value will be restricted. That is, what are the implications of certain forms of environmental damage for the lives of individuals? Concentrating on what individuals may be willing to pay to avoid certain losses does not provide information about the ways in which the genuine opportunities open to individuals may change should certain hazards be realized.

### A Capability Approach to Risk Determination

In a capability approach, the consequences of hazardous scenarios are conceptualized and assessed in terms of changes in capabilities. Risk is then defined as the probability that capabilities will be reduced (Murphy and Gardoni 2006). Conceptualizing the consequences component of risk from a capability approach overcomes the limitations with both the resource and utilitarian framework. It focuses the attention of risk analysts directly on how a hazardous scenario will affect the welfare of individuals within a community, insofar as consequences are defined in terms of the change in the genuine opportunities of individuals to achieve valuable states and activities. It thus avoids the mistake of focusing on commodities, which are instrumental means but not valuable ends in themselves. Furthermore, the capability approach can account for differences in interpersonal conversion rates in using resources to achieve valuable functionings; the capability approach focuses directly on the freedom to achieve valuable functionings itself. The capability approach also does not define the consequences of hazards in terms of what individuals would be willing to pay. It thus avoids the commodification of costs like the loss of a human life and problems with the lack of information about, concern for, or ability to avoid certain risks plaguing the utilitarian analysis. What individuals would be willing to pay to avoid certain consequences plays no role in determining what the consequences of a hazardous scenario might be. Instead, the impact of a hazard on the genuine opportunities of individuals is the focal point of concern. The capability approach can account for the broader indirect impact that a hazard might have, because built into the framework is the recognition of the fact that both resources and the social and material structure of a community profoundly affect the opportunities open to individuals. Concern with the freedom of individuals to achieve valuable functionings invites questions about whether and in what ways the resources of individuals or the social and material structure of a community have changed. Finally, concentrating on capabilities, rather than achieved functionings, reflects a respect for the commitment of liberal governments to ensure that a range of options of ways of living are open to individuals, instead of promoting a particular way of living itself.

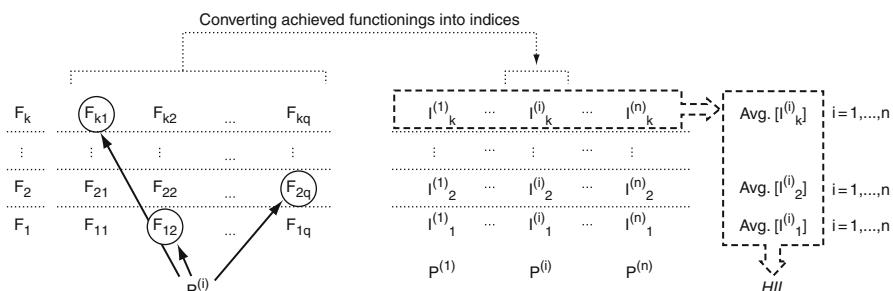
To assess the impact of a hazard, it is necessary to operationalize the capability approach to risk. Initial efforts to operationalize the capability approach built on the framework are provided by the *HDI* for development, which assesses achieved functionings (Gardoni and Murphy 2009). For the purposes of quantifying the consequences of risk, a Hazard Impact Index *HII* was developed (Gardoni and Murphy 2009). The *HII* is constructed in four steps. The first step involves selecting the capabilities relevant for the kind of risk in question; that is, it is necessary to identify which capabilities will provide an accurate picture of the impact of a hazard. Gardoni and Murphy (2009) have argued for a set of criteria that any argument for the capabilities to be used in risk analysis must satisfy. Briefly, a given capability is relevant for risk analysis insofar as it reflects the underlying values and concerns with risk. To ensure that a capability risk analysis is

both comprehensive and, at the same time, practically possible to perform, the minimum number of capabilities should be selected and the capabilities chosen should provide information not available from given capabilities (e.g., because there is a correlation between the enjoyment of two capabilities). Gardoni and Murphy also proposed a preliminary list of capabilities to be used in risk analysis conducted with respect to natural hazards.

Indicators must then be chosen for the selected capabilities. In the context of risk, appropriate indicators must be shown to track the impact on the capability of interest of a hazardous scenario (Gardoni and Murphy 2010). Such indicators will allow risk analysts to assess the impact of a hazardous scenario on capabilities, and so assess the consequence component of risk. Following the model provided by the *HDI* indicators were taken to track the change in the level of achievement of each associated functioning. Each indicator must be converted into an index that ranges from 0 (minimum achievement) and 1 (maximum achievement) through a process of scaling analogous to that used in the *HDI*. Finally, all indices are combined through an averaging process to create an aggregate measure of achievement. It is also noted that in the context of risk analysis the values of the indicators cannot be simply measured but need to be predicted since the subject of risk analysis is future events. The *HII* gives a snapshot of the predicted well-being after a hazard. A comparison between the *HII* and the actual well-being before a hazard (also measured using a capability approach) provides a measure of the impact of the hazard.

Figure 39.1 illustrates this formulation of the *HII*.  $F_1, F_2, \dots, F_k$  on the left column are the functionings under consideration. Each functioning can be achieved at different levels 1 through  $q$ . For example, functioning  $F_k$  might be achieved at levels  $F_{k1}, F_{k2}, \dots, F_{kq}$ . As shown in Fig. 39.1, an individual  $P^{(i)}$  might achieve  $F_1$  at level  $F_{12}$ ,  $F_2$  at level  $F_{2q}$  and so on, up to  $F_k$  at level  $F_{k1}$ . Each level of achieved functioning for individual  $P^{(i)}$  is then converted into a corresponding index, for a total of  $k$  indices:  $I_1^{(i)}, I_2^{(i)}, \dots, I_k^{(i)}$ . The same process is repeated for all the considered individuals  $P^{(1)}, \dots, P^{(n)}$ . An average of the indices over all individuals is then computed ( $\text{Avg.}[I_k^{(i)}]$ ,  $i = 1, \dots, n$ ) for each functioning. Finally the *HII* is computed by combining all the averages.

One limitation with this construction of the *HII* is that it assesses the level of achievement of functionings, not capabilities. For assessing the least advantaged in society and for assessing the impact of a natural hazard (e.g., flood, earthquake, or hurricane) in the emergency phase



**Fig. 39.1**

Illustration of the current formulation of the *HII* and *HDI* (Published previously as Fig. 1, in Murphy and Gardoni 2010, with kind permission from Taylor & Francis Ltd., <http://www.informaworld.com>)

(the time that immediately follows a medium or large disaster) information about achieved functionings may be sufficient for assessing capabilities. This is because we can typically assume that the functionings frequently used in these contexts, which are basic functionings, will be automatically chosen if individuals have the genuine opportunity to do so. Examples of basic capabilities in Murphy and Gardoni's (2010) sense include the capability to have adequate shelter, avoid injuries, and be adequately nourished. Murphy and Gardoni's use of the notion 'basic capabilities' departs from both Nussbaum and Sen. Nussbaum defines basic capabilities as "the innate equipment of individuals that is the necessary basis for developing the more advanced capabilities, and a ground of moral concern." (2000b, p. 84). Sen defines basic capabilities as capturing "the ability to satisfy certain crucially important functionings up to certain minimally adequate levels." (1993, p. 40). By contrast, we delimit basic capabilities in terms of those functionings that, with rare exceptions, individuals will choose to achieve if they have a genuine opportunity to do so. In the case of basic capabilities, the levels of achieved functionings thus provide an accurate picture of the capabilities. For example, we can reasonably assume that if some individuals in the emergency phase in the aftermath of a hazard are not sheltered this is because they lack the capability to be so. However, when we gage the medium- and long-term impacts of a hazard on capabilities, an assessment of functionings achievement is typically an inaccurate proxy. This is because the medium- and long-term impacts of large hazards, or the impact of smaller hazards, affect non-basic functionings (or basic capabilities in a more subtle manner) that we cannot safely assume will be chosen given an opportunity to do so.

In addition, as discussed earlier, in practice only vectors of functionings are open to an individual and choosing to achieve one functioning level might preclude the opportunity to chose a certain level of a different functioning. That is, only certain combinations of functionings are open to an individual to achieve. However, the original formulation treated each functioning in isolation and did not account for how different functionings might interact in fact. (Murphy and Gardoni 2010). Furthermore, the original formulation assessed the average impact on levels of functionings achievement across a population. However, an average impact can mask important variations among a population. The same average may reflect a case in which all individuals had roughly the same impact, or a case in which some individuals were not impacted but a portion of the population was severely impacted. The variation in impact may correlate with certain subgroups within a population, though it need not do so. Finally, in evaluating the capability of an individual we need to consider two dimensions captured by this approach: the *quality of the possible vectors*, and the *extent of the freedom* that an individual has in choosing different vectors. In evaluating capability, we need to capture both the quality of the options open to an individual as well as the quantity, or range of vectors, he or she has a genuine opportunity to achieve. The quality of the possible vectors provides information about the levels and kinds of functionings that can be achieved, and, more generally, of the available opportunities. The range of vectors provides information about the scope of freedom.

For this reason, Murphy and Gardoni (2010) proposed a method for determining vectors of capabilities of an individual. The fundamental idea is that functioning achievements of other people says something about the choices that are genuinely open to an individual in the same group. Knowing that other individuals have chosen to realize a certain opportunity is an indication that such opportunity actually exists and allows us to distinguish between a situation where an individual chooses not to actually take advantage of an opportunity versus a situation in which such an opportunity does not exist. This method also accounts for

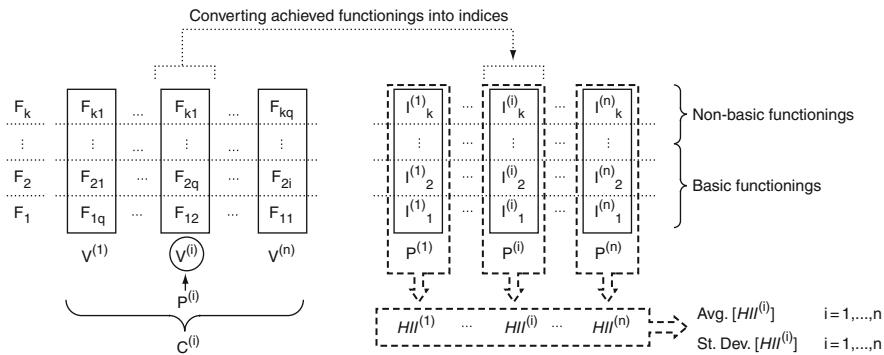


Fig. 39.2

Illustration of the new formulation of the  $HII$  (Published previously as Fig. 2, in Murphy and Gardoni 2010, with kind permission from Taylor & Francis Ltd., <http://www.informaworld.com>)

vectors, rather than isolated functionings, by computing the impact of a hazard at the individual level and then determining the achieved functionings across individuals.

Figure 39.2 illustrates the formulation proposed by Murphy and Gardoni (2010). An individual  $P^{(i)}$  can only choose among a set of vectors of achieved functionings,  $V^{(1)}, \dots, V^{(n)}$ . For example,  $P^{(i)}$  can choose  $V^{(i)}$  which means choosing  $F_{12}, F_{2q}, \dots, F_{k1}$ . After transforming the achieved functionings into indices, we need to compute the  $HII^{(i)}$  (or the  $HDI^{(i)}$  if this approach is applied to human development) for individual  $P^{(i)}$  and only then determine the statistics (average and standard deviation) of the  $HII^{(i)}$  across all individuals, indicated in Fig. 39.2 as  $\text{Avg.}[HII^{(i)}]$  and  $\text{St. Dev.}[HII^{(i)}]$ , respectively. In this formulation,  $\text{Avg.}[HII^{(i)}]$  tells us about the quality of the possible options, while the  $\text{St. Dev.}[HII^{(i)}]$  captures the extent of freedom.

## Risk Evaluation

As its name implies, risk evaluation is the process of assessing the information provided through a risk determination. Of particular interest is whether a risk is acceptable or not. Below we survey two common methods for judging the acceptability of risks and their corresponding limitations. A third approach, cost-benefit analysis, is considered in the next section on risk management. We then present the capability approach to acceptable risk.

### Limitations with Common Frameworks to Risk Evaluation

One approach to determining the acceptability of a risk is to consider whether the public finds a given risk acceptable (Hunter and Fewtrell 2001). Underlying this approach is a commitment to democratic governance, where democracy is taken to imply a form of majority rule. In a democracy, policies and decisions are justified insofar as they express the will of the people. However, there are two important limitations with using public judgment as the basis for determining the acceptability of risks. First, as noted earlier, the public often has limited, and sometimes inaccurate, knowledge about risks; thus judgments will not be based on correct assessments. Furthermore, the public may be indifferent about risks they face. All of these factors undermine the confidence we can have in the assessments of the acceptability of a given

risk the public may make. Second, the public judgment standard for acceptable risk mistakenly equates democracy with majority rule. However, it is widely acknowledged by liberal democratic theorists that individual rights are at the core of democracy and constrain what the majority may want to do. Furthermore, democratic processes should not take preferences or views of citizens as a given, but rather aim to help develop an informed citizenry. Thus, democracy does not require that the initial, often uninformed, preferences of citizens be the basis for evaluating risks a community faces.

An alternative approach is to leave judgments about the acceptability of risks to experts. Professional risk analysts have the most complete information about the character of risks and the knowledge to understand how to interpret the results from a risk analysis. Often risk analysts use as the basis for their judgments their past experience, the standards of their profession, or the preferences and desires of their client (Fischhoff et al. 1981). However, in practice experts may be biased in their evaluation of the acceptability of risks, failing to take into consideration aspects of risk that matter to the general public or how risks impact society as a whole, as opposed to their client in particular. Finally, the criteria experts use to judge a risk as acceptable are frequently implicit, making it difficult to scrutinize or evaluate the justifiability of a particular decision.

### A Capability Approach to Risk Evaluation

In addition to providing an alternative characterization of the consequences of a hazardous scenario, the capability approach provides resources for a novel way of conceptualizing the criteria in terms of which a risk should be judged acceptable. Motivating this approach is the key claim that justice requires that individuals enjoy a certain threshold level of capability (Nussbaum 2000b). Judgments about the acceptability of a risk, then, should take into account whether or not individuals will continue to enjoy threshold levels of capabilities in the aftermath of a hazard (Murphy and Gardoni 2008).

There are two thresholds that are relevant in this context. The first is the acceptable threshold. This threshold specifies what minimum level of capabilities it is acceptable for individuals to enjoy over any period of time; this is the level below which capabilities should not fall ideally. In the aftermath of a hazard, especially the immediate aftermath, it may not be practically possible for all individuals to enjoy the acceptable level of capabilities. A framework for risk evaluation should recognize this possibility, while at the same time setting limits on what is permissible even in the immediate aftermath of a hazard. This is where the second threshold comes in. The tolerability threshold is lower than the acceptable threshold. It specifies what level of capabilities is tolerable in the immediate aftermath of a hazard, provided this lower level is temporary, reversible, and does not fall below a lower tolerable limit. The tolerability threshold reflects the moral necessity of avoiding permanent damage to the well-being of individuals (Murphy and Gardoni 2008).

To account for the role of public deliberation, Nussbaum (2000b) and Sen (2009) rightly note, is an important component of the capability approach. In a capability approach to acceptable risk (Murphy and Gardoni 2008), the precise specification of the acceptable and tolerable threshold can be the product of internal democratic processes. The goal of such public processes is to articulate standards that establish realistically ambitious levels of acceptable and tolerable capabilities, given a society's actual conditions. Because we are dealing with risks, in which the precise consequences of a hazard are unknown, the specification of the thresholds of acceptable risk must also take into account probability. That is, in addition to specifying the

level of capabilities judged acceptable, the acceptable threshold must also specify the probability that capabilities will not fall below the acceptable threshold which must be minimally met for a risk to be judged acceptable. The same consideration is true of the tolerability threshold.

In risk evaluation, then, we assess the probability that the level of capabilities across a population is below the thresholds of acceptable and tolerable risk. To take into consideration questions of justice in the distribution of risks, a process of disaggregation may be used. This involves determining the probability distribution (as a measure of the likelihood) of the capabilities of subgroups within a population. The probability that the level of capabilities for various subgroups is below the thresholds of acceptable and tolerable risk may then be assessed.

The capability approach to acceptable risk avoids the limitations with relying either only on public judgment or on expert judgment. It finds a central role for the public to play in shaping the criteria of acceptable risk in setting the thresholds of acceptability. At the same time, judgments about the acceptability of particular risks can be done by experts, whose evaluation will be shaped by the standards the public sets.

## Risk Management

Communities and individuals want to know what risks they face, and to make judgments about whether such risks are acceptable, in part so that they can effectively manage risks. The information provided by risk determination and risk evaluation provides a foundation for ensuring that policies about risk are formulated in a well-informed manner. Such policies must take into account the limited resources available to governments and agencies as well as the other competing priorities that governments must address. Risk management strategies provide a framework for aiding policy makers, who are generally not experts in risk analysis, in deciding which risks should be prioritized for mitigation action and which mitigation strategies should be taken. Below we first discuss one prominent risk management framework, cost-benefit analysis and its limitations. We then consider a capability approach to risk management.

### Limitations with Common Frameworks to Risk Management

A common method of evaluating and managing risks is cost-benefit analysis (CBA). CBA is a broadly utilitarian approach in which two or more options for public policy decision making are compared on the basis of their respective advantages and disadvantages. The advantages and disadvantages are defined in terms of the consequences. All consequences (which may include economic costs, risks of death, environmental harm) are assigned a numerical, typically monetary, value. The monetary value is assessed based on the preferences of individuals, where such preferences are measured based on the amount of money an individual would be willing to pay to avoid certain risks or the amount of money he or she would agree to be paid to be exposed to certain risks. Market information or valuation surveys provide the basis for monetary figures. After quantifying the relative advantages and disadvantages, advantages are weighed against the disadvantages numerically. The option with the greatest net benefit (taking into account benefits minus costs) is then selected (Sunstein 2005; Hansson 2007a). There are two strengths of CBA, which explains in part why it is favored by policy makers. First, it provides a straightforward decision-making strategy for determining which risks to address and in what way to address them. Second, it takes as a priority the efficient allocation of scarce resources.

However, CBA has also been the subject of a number of criticisms, which parallel those raised to a utility approach to risk determination. To briefly review, first, as noted earlier, many critics of cost-benefit analysis object to allowing the market to establish a distribution of risks on the basis of what individuals are willing to pay or be paid. For critics, this does not recognize and cannot guarantee the obligations that both employers and communities have to guard against exposing employees or individuals to certain kinds of risks. Second, relying on market mechanisms to quantify risks is problematic because it assumes that individuals have full knowledge about the risks they face when they determine what they would be willing to be paid to be exposed to certain risks and that they base their decision on their assessment of how they value what will be put at risk, such as their lives. However, neither assumption is warranted in practice. As noted above, individuals are often not fully knowledgeable about the risks they face. Furthermore, choices about risk exposure are often a function of factors, such as limited income or employment opportunities, that do not reflect how individuals value certain goods. Third, CBA fails to sufficiently recognize that the protection of life is not simply a question to which a certain cost should be assigned, but reflects a principle or ideal to which societies should be committed. Relatedly, it is problematic to assess a monetary value to the loss of human life or a monetary value to other noneconomic goods. Fourth, there are important issues of fairness and equality in the distribution of risks which the CBA cannot take into account, given its emphasis on the aggregate costs and benefits (Asveld and Roeser 2009, esp. the chapters by Hansson, Cranor, and MacLean).

### A Capability Approach to Risk Management

In a Capability Approach to risk management, the overarching aim is to protect and promote individuals' capabilities (Murphy and Gardoni 2007). Given this aim, the first step in evaluating which risks to address is to determine whether a given risk is acceptable or tolerable. Priority should be given to risks judged intolerable, followed by a consideration of risks judged unacceptable.

When evaluating a range of policy options to address and mitigate unacceptable or intolerable risks, policy makers should determine which policies are viable options. Viable options should be reasonably anticipated to bring the predicted level of capabilities above the tolerable or acceptable thresholds, respectively. This can be achieved by reducing the probability of occurrence of a given hazard and/or its expected impact. Because resources are limited and risk management is not the only objective to pursue, choices among viable options need to be made. In a Capability Approach, we select from among viable options on the basis of the likely affectability of relative policies. That is, from among policies considered, we can select a policy to pursue based on the expected dollar per unit change in the impact of a hazard. Monetary quantification can be used as a measure of the cost of a given policy; its effectiveness is then conceptualized in terms of the increase or decrease in capability levels.

The capability approach to risk management avoids some of the main criticisms leveled against CBA. In a capability approach the thresholds that set the criteria for acceptable and tolerable risk reflect the judgment that justice requires that individuals not be exposed to certain kinds of risks and that society has an obligation to ensure that individuals are protected from such risks. Because the quantification of the impact of risks and of mitigation strategies is defined and assessed in terms of capabilities, the Capability Approach does not involve quantifying the value of a human life or commodifying diverse kinds of goods. At the same

time, assessing policies on the basis of their likely affectability ensures that consideration is given to the fact that resources are scarce and that it is important to ensure that public resources are used in an efficient manner.

## Use of Risk Theory in the Capability Approach

---

The concept of risk has been used to enrich our understanding of capability. In their work on assessing advantage and disadvantage from a capability approach, philosophers Jonathan Wolff and Avner de-Shalit (2007) have argued that it is important to consider the security of any functioning achievement when assessing the capability of an individual. An individual has a genuine opportunity to achieve a given functioning only if that achievement is one that can be sustained; it is not an achievement that is at undue risk. A functioning achievement is at undue risk, in their view, insofar as its achievement is temporarily and involuntarily put in jeopardy by the pursuit of other functionings achievement. In other words, a functioning achievement is insecure when an individual is “forced to take risks that in one way or another are bigger than others are being exposed to or take” (Wolff and de-Shalit 2007). To illustrate, the achievement of being sheltered may be insecure and at risk insofar as an individual has an uncertain income stream stemming from irregular employment that results in unstable ability to pay rent. Alternately, an individual may be forced to constantly put her bodily integrity at risk, insofar as the only available sources of employment are dangerous occupations (Wolff and de-Shalit 2007).

The inclusion of the idea of secure functionings achievement is becoming widely accepted and adopted in capability analyses. Capability is now understood to represent the genuine opportunity an individual has to achieve a particular functioning in a secure manner. Thus, concern with the functionings achievements being put at risk in an involuntary or disproportionate manner is a central issue. Wolff and de-Shalit suggest statistical information can provide some insight into the security of functionings achievement. In their view, we can assess the security of functionings achievement by looking at trends over time for specified subgroups. In their words, (2007, pp. 116–117),

- ▶ The fact that I have a job today says nothing about whether it is a day's casual work or a sinecure for life. However, taking a wider view and looking at the individual's social circumstances or context immediately provides more information. Indeed, statistics will provide much of what we need. Imagine that among certain groups – perhaps the young, recent immigrants, or the low paid – there is a high degree of mobility in employment or housing, with those moving jobs or homes also experiencing periods of unemployment or homelessness. This, then, gives a *prima facie* reason to believe these functionings are not achieved securely by people within these groups... In general, then, although individual functioning is not an indicator for the degree of security, statistics often can be.

There is a second contribution that risk theory makes to the capability approach. The method developed to assess capabilities in the context of risk analysis (Murphy and Gardoni 2010) and illustrated in Fig. 39.2 can also be applied to assess capabilities in its original context of development economics and in other applications. As noted above, initial efforts to operationalize the capability approach built on the model provided by the *HDI* for development, which concentrated on assessing achieved functionings. Furthermore, the

original formulation treated each functioning achievement in isolation, while in practice an individual  $P^{(i)}$  can only choose among a set of vectors of achieved functionings.

## Further Research

---

There are two main areas of further research for a capability approach to risk analysis that we concentrate on in this section. The first centers on the issues to address to operationalize the capability approach to risk. The second centers on the place of a concern with capabilities in an overall perspective on risk evaluation and management.

The capability approach developed by Sen is deliberately underspecified. Sen (1992, 1993) refrains from identifying the capabilities that should be used in any given evaluative exercise for two reasons. The first is that the capabilities that should be considered will always be to some extent evaluative exercise specific; that is, the capabilities of interest when considering development are not necessarily the capabilities of interest when assessing risk. The second reflects Sen's commitment to identifying valuable capabilities via democratic procedures, and in particular democratic deliberation. These two commitments of Sen's give rise to a series of theoretical questions for a capability approach to risk. First, there is the issue of how evaluative exercise specific the selection of capabilities should be. For example, should the capabilities selected vary according to the kind of risk being assessed? The work done to date developing a capability approach to risk has been conducted in the context of the risks posed by natural hazards. Are the capabilities relevant for technological risks, for example, the same as those relevant for natural hazards? On what basis do we answer this question? Second, if Sen is correct about the role of public deliberation, in the context of risk there is a need to incorporate a place for public deliberation in the selection or evaluation of capabilities and a method for such deliberation to become possible.

Making the capability approach operationalizable also requires further research into a cluster of questions surrounding the quantification of consequences. Indicators must fulfill at least two criteria (Gardoni and Murphy 2010). They must be representative of a given capability and be intuitively plausible. The latter criterion is required to facilitate the adoption and successful application of a capability approach in public policy discussion and decision making about risk. Indicators in the context of risk are designed to track changes in capabilities, so that we can understand the impact of a given hazardous scenario on a community. One question concerns which indicators will be both representative of a capability and provide information about the change in capabilities due to a hazard.

Another issue is that, unlike many current applications of the capability approach, indicators in the context of risk must be *predicted*. With risks we are always dealing with future possible states, not analyzing existing states as is the case with many current applications of the capability approach. The basis on which an accurate and credible prediction of changes in capabilities can be made must be determined and, in particular, whether current methods for quantifying risks can be used for this purpose or whether novel methods for prediction must be developed. One method (Gardoni and Murphy 2010; Murphy and Gardoni 2010) is to forecast the impact of future hazards by first assessing the impact of past disasters using available data. The impacts of past disasters can then be used to estimate the likelihood of each potential outcome of  $HII$  for a given hazard type,  $H$ , of magnitude,  $M$ . Mathematically, such likelihood is expressed as the conditional Probability Density Function (*PDF*),  $P(HII|H, M)$ , of  $HII$  for given  $H$  and  $M$ . Using the Total Probability Rule (Ang and Tang 2007), we can then estimate the

probability of future societal impacts accounting for their likelihood of occurrence integrating out  $H$  and  $M$ , as

$$P(HII) = \int P(HII|H, M)P(H, M)dHdM$$

where  $P(H, M)$  is the joint PDF of  $(H, M)$ , which can be estimated for typical natural hazards based on for example meteorological or seismological considerations for hurricanes and earthquakes, respectively.

Further research is also required to determine whether, and to what extent, indicators for selected capabilities must be either time-dependent or risk-dependent. With respect to time-dependence, the issue is whether in trying to understand the long-term versus immediate impact of a hazard the indicators must change.

The second general topic for further research concerns the place of capabilities within a comprehensive framework of risk evaluation and risk management. One specific question for further research is the place of a consideration of capabilities among the additional, diverse moral considerations that a judgment of the acceptability of a risk must take into account. The capability approach to risk developed to date provides an account of how the capabilities dimension of risk should be evaluated, which reflects the fact that capabilities are constitutive dimensions of individual well-being and that justice requires a minimum level of capabilities be guaranteed for all individuals. However, the consequences of a hazardous scenario are not the only factor that is relevant for risk evaluation purposes. It is well-established that the public is concerned about and distinguishes among risks on the basis of the source of a risk, that is, how a risk was brought about or created (Wolff 2006; Murphy and Gardoni 2011). The public distinguishes among risks on the basis of whether they were created through negligence, recklessness, and non-culpable behavior. Risks are also evaluated differently depending on whether they were voluntarily taken or not. The distribution of benefits associated with taking risks and the potential harm that may be realized should a hazardous scenario occur is morally salient (Fischhoff et al. 1981; Slovic 2000; Asveld and Roeser 2009). One question for further research then is: when evaluating a risk, on what basis should we weigh or take into consideration the expected impact of a given risk on capabilities relative to the other moral factors that matter? A second question centers on the broader implications of a capability approach for issues in risk management. That is, if we conceptualize risks from a capability perspective, how should that inform the principles that should guide the creation and regulation of risk? Consider technological risks. Should, for example, new technologies be permitted only insofar as it can be shown that such technologies will enhance human capabilities? Should the design of technologies be guided by a concern for capabilities? There are a number of authors working on extending the capability approach to areas like design ethics (Oosterlaken 2009), information technology (Johnstone 2007; Wresch 2007; Zheng 2007; Coeckelbergh 2010b) and AI technology in health care (Coeckelbergh 2010a). Another question is whether it is necessary to adopt a capability approach to design or to technology insofar as one adopts a capability approach to risk.

## Conclusion

This chapter has provided an overview of the reciprocal contributions of a capability approach and risk theory. A capability approach can enrich risk theory in three ways: (1) The consequences component of risk can be conceptualized and assessed in terms of the impact of

a hazardous scenario on capabilities; (2) Judgments of the acceptability of risks can be made on the basis of whether certain threshold levels of capabilities are maintained; (3) The primary aim of risk management can be to protect and promote individuals' capabilities, by prioritizing unacceptable or intolerable risks and by selecting policy options on the basis of expected dollar per unit change in the impact on capabilities. On the other hand, risk theory has enriched the capability approach by (1) drawing attention to an important dimension of capability, namely security; and (2) providing resources for developing a novel way to assess capabilities instead of functioning. Further work is needed to make the capability approach to risk theory operationalizable and to clarify the place of a concern for capabilities in an overall perspective on risk evaluation and management.

## Acknowledgment

This research was supported primarily by the Science, Technology, and Society Program of the National Science Foundation Grant (STS 0926025). Opinions and findings presented are those of the authors and do not necessarily reflect the views of the sponsor.

## References

- Anderson E (1988) Values, risks, and market norms. *Philos Public Aff* 17(1):54–65
- Ang AH-S, Tang WH (2007) Probability concepts in engineering: emphasis on applications to civil and environmental engineering. Wiley, New York
- Asveld L, Roeser S (2009) The ethics of technological risk. Earthscan, London
- Bedford T, Cooke R (2001) Probabilistic risk analysis: foundations and methods. Cambridge University Press, Cambridge
- Coeckelbergh M (2010a) Health care, capabilities, and AI assistive technologies. *Ethical Theory Moral Pract* 13:181–190
- Coeckelbergh M (2010b) Human development or human enhancement? A methodological reflection on capabilities and the evaluation of information technologies. *Ethics Inf Technol*. doi:10.1007/s10676-010-9231-9
- Crocker D (2008) The ethics of human development. Cambridge University Press, New York
- Cullen A, Small MJ (2004) Uncertain risk management choices in risk analysis and society. In: McDaniels T, Small MJ (eds) Risk analysis and society: an interdisciplinary characterization of the field. Cambridge University Press, Cambridge, pp 163–212
- Fischhoff B, Lichtenstein S, Slovic P, Derby SL, Keeney RL (1981) Acceptable risk. Cambridge University Press, Cambridge
- Fukuda-Parr S, Kumar AKS (eds) (2003) Readings in human development. Oxford University Press, Oxford
- Gardoni P, Murphy C (2009) A capabilities-based approach to measuring the societal impacts of natural and man-made hazards. *Nat Hazard Rev* 10(2):23–37
- Gardoni P, Murphy C (2010) Gauging the societal impacts of natural disasters using a capabilities based approach. *Disasters J Disaster Stud Policy Manag* 34(3):619–636
- Haines YY (2004) Risk modeling, assessment, and management, Systems Engineering and Management. Wiley, Hoboken
- Hansson SO (2007a) Philosophical problems in cost-benefit analysis. *Econ Philos* 23:163–183
- Hansson SO (2007b) Risk. Stanford encyclopedia of philosophy. <http://plato.stanford.edu/entries/risk/>. Accessed 23 March 2010
- Hunter PR, Fewtrell L (2001) Acceptable risk. In: Fewtrell L, Bartram J (eds) Water quality: guidelines, standards, and health. IWA, London, pp 207–227
- Jahan S (2003) Evolution of the human development index. In: Fukuda-Parr S, Shiva Kumar AK (eds) Readings in human Development. Oxford University Press, Oxford, pp 128–139
- Johnstone J (2007) Technology as empowerment: a capability approach to computer ethics. *Ethics Inf Technol* 9:73–87
- Kaplan S, Garrick BJ (1981) On the quantitative definition of risk. *Risk Anal* 1:11–27
- May P (2001) Organizational and societal consequences for performance-based earthquake engineering. *PEER 2001/04, Pacific Earthquake Engineering*

- Research Center, College of Engineering, University of California–Berkeley, Berkeley
- Murphy C, Gardoni P (2006) The role of society in engineering risk analysis: a capabilities based approach. *Risk Anal* 26(4):1085–1095
- Murphy C, Gardoni P (2007) Determining public policy and resource allocation priorities for mitigating natural hazards: a capabilities-based approach. *Sci Eng Ethics* 13(4):489–504
- Murphy C, Gardoni P (2008) The acceptability and the tolerability of risks: a capabilities-based approach. *Sci Eng Ethics* 14(1):77–92
- Murphy C, Gardoni P (2010) Assessing capability instead of achieved functionings in risk analysis. *J Risk Res* 13(2):137–147
- Murphy C, Gardoni P (2011) Evaluating the source of the risks associated with natural events. *Res Publica* 17(2):125–140
- Nussbaum M (2000a) Aristotle, politics, and human capabilities: a response to Antony, Arneson, Charlesworth, and Mulgan. *Ethics* 111(1):102–140
- Nussbaum M (2000b) Woman and human development: the capabilities approach. Cambridge University Press, Cambridge
- Nussbaum M (2001) Adaptive preferences and women's options. *Econ Philos* 17:67–88
- Oosterlaken I (2009) Design for development: a capability approach. *Des Issues* 25(4):91–102
- Paté-Cornell ME (1996) Uncertainties in risk analysis: six levels of treatment. *Reliability Eng Sys Safety* 54:95–111
- Raworth K, Stewart D (2003) Critiques of the human development index: a review. In: Fukuda-Parr S, Shiva Kumar AK (eds) Readings in human development. Oxford University Press, Oxford, pp 140–152
- Robeyns I (2005) The capability approach: a theoretical survey. *J Hum Dev* 6(1):93–114
- Robeyns I (2006) The capability approach in practice. *J Polit Philos* 14(3):351–376
- Rowe WD (1980) Risk assessment: theoretical approaches and methodological problems. In: Conrad J (ed) Society, technology, and risk assessment. Academic, New York, pp 3–29
- Sen A (1989) Development as capabilities expansion. *J Dev Plan* 19:41–58
- Sen A (1992) Inequality reexamined. Harvard University Press, Cambridge
- Sen A (1993) Capability and well-being. In: Nussbaum M, Sen A (eds) The quality of life. Clarendon, Oxford, pp 30–53
- Sen A (1999a) Commodities and capabilities. Oxford University Press, Oxford
- Sen A (1999b) Development as freedom. Anchor Books, New York
- Sen A (2009) The idea of justice. Belknap Press of Harvard University Press, Cambridge
- Slovic P (1987) Perception of risk. *Science* 236:280–285
- Slovic P (2000) The perception of risk. Earthscan, London
- Sunstein C (2005) Cost-benefit analysis and the environment. *Ethics* 115:251–285
- UNDP (United Nations Development Program) (2000) Human development report 2000. Oxford University Press, New York
- UNDP (United Nations Development Program) (2007) Human development report 2007. Oxford University Press, New York
- Vose D (2000) Risk analysis: a quantitative guide. Wiley, New York
- Wolff J (2006) Risk, fear, blame, shame and the regulation of public safety. *Econ Philos* 22:409–427
- Wolff J, de-Shalit A (2007) Disadvantage. Oxford University Press, Oxford
- Wresch W (2007) 500 million missing web sites: Amartya Sen's capability approach and measures of technological deprivation in developing countries. In: Rooksby E, Weckert J (eds) Information technology and social justice. Information Science, Hershey
- Zheng Y (2007) Exploring the value of the capability approach for E-development. Paper presented at the 9th international conference on social implications of computers in developing countries, Sao Paulo



# Part 6

## Risk in Society



# 40 Sociology of Risk

Rolf Lidskog<sup>1</sup> · Göran Sundqvist<sup>2</sup>

<sup>1</sup>Örebro University, Örebro, Sweden

<sup>2</sup>University of Oslo, Oslo, Norway

<b>Introduction .....</b>	<b>1002</b>
<b>History .....</b>	<b>1004</b>
What Is Sociology? .....	1004
Risk in Sociology: From Social Problems to Risk .....	1005
Sociology Explaining Public Misperceptions of Risk .....	1006
Sociology Explaining Amplifications of Risk .....	1007
Sociology Explaining Risk .....	1008
<b>Current Research .....</b>	<b>1010</b>
Important Theories .....	1010
Mary Douglas: Purity and Danger .....	1010
Ulrich Beck: The Risk Society and Reflexive Modernization .....	1012
Niklas Luhmann: System Theory and Risk .....	1014
Thematic Areas .....	1016
Organizational Risk: From Risk Analysis to Risk Governance .....	1016
Public Trust: The Relation Between Experts and Laypeople .....	1017
Risk and Democracy: The Importance of Framing .....	1019
Produced Risk: Beyond Realism and Constructivism .....	1020
Governmentality: Toward an Individualized Risk Management .....	1022
<b>Further Research .....</b>	<b>1024</b>

**Abstract:** Risk is a relatively new object of sociological research, but this research field has grown rapidly over the last three decades. This chapter argues that the task of sociology is to contribute to risk research by emphasizing that risks are always situated in a social context and are necessarily connected to actors' activities. Thus, sociology opposes the reification of risks, where risks are lifted out their social context and are dealt with as something uninfluenced by activities, technologies, and instruments that serve to map them. The chapter comprises four sections, the first being a general introduction. The next section presents a historical perspective on sociology and risk. It starts by briefly describing what sociology is, followed by a discussion of how the concept of risk becomes an object of sociological thought. The section ends by presenting and discussing three different sociological perspectives on risk, all providing different contributions. The third section focuses on current strands of sociological risk research. It starts by giving an overview of three different sociological approaches to risk: Mary Douglas's cultural theory of social order, Ulrich Beck's theory of reflexive modernization and the risk society, and Niklas Luhmann's system theory. All three approaches conceptualize risk differently and make different contributions to the sociological study of risk. The review of these theories is followed by a presentation of five central discussions within the sociology of risk: risk governance, public trust, democracy and risk, the realism–constructivism debate, and governmentality and risk. Finally, the fourth section briefly presents some areas in need of further sociological research.

## Introduction

---

Over the last few decades, we have witnessed an explosion of risk management practices across a wide range of organizational contexts, such as environment, health, food, crime, media, and traffic (Lupton 1999; Tulloch 1999; Hutter and Power 2005; Hughes et al. 2006; Kemshall 2006; Taylor-Gooby and Zinn 2006; Renn 2008; Reith 2009). All organizations have to relate to risks in their environment. These risks are not only connected with industrial activities, such as harmful substances or technical artifacts. Instead, a growing number of risks concern how actors act upon what they see as risks associated with an organization. Public relations, risk communication, and participatory approaches to risk management have emerged as means to handle diverging interests in society; not least public perceptions could be a source of risk in the sense that these perceptions could pose a threat to the legitimacy and stability of existing ways of managing risk (Power 2007, p. 21). Thus, risk management focuses on how organizations deal not only with the technical calculation of risks, but also with the actors they perceive as possible threats and potential risks to the stability of the organization. Managing such processes is a matter not only of rules for how we should mitigate or accept certain environmental hazards or health risks, but also of rules regarding the process itself, and of activities that target the understanding of risk and deal with public opinion and perceptions concerning it.

This development implies that risk management is no longer limited to a specific sector dealing with certain kinds of risks (such as nuclear power, the chemical industry, and road transport). Risk management has instead become an integral part of managerial language and organizational activities, and all organizations – private companies, governmental agencies, interest organizations, and nongovernmental organizations – have to deal with risks and have made risk management an important rationale for their activities. Not only organizations, but also citizens must include risk thinking when organizing their social world

(Tulloch and Lupton 2003; Höijer et al. 2006). Previous certainties – social forms such as nation state, class, ethnicity, traditional family structures, and gender roles – which people use to map out their future are now eroding and citizens have to navigate their lives without them (Beck 2002, p. 22). There is not only public concern about new technologies, but also about how to organize social life and who and what to trust in an uncertain world. This has led researchers to claim that we today face a grand narrative of risk and risk management at the global level (Power 2007, p. viii). Thus, society has no option but to organize itself in the face of risk (Lidskog et al. 2005). Assessing, managing, and communicating risk has become a veritable industry.

This progression from risks associated with certain industrial activities to risks associated with individual and organizational behavior has led to a strong call for sociological analysis. However, scientific development is not only a reflection of changing societal conditions; it is also a driving force of this change. Social theorists have claimed that we today live in a risk society (Beck 1992), a culture of fear (Furedi 2002), and in a social climate that fosters insecurity, fear, and risk (Giddens 1990; Bauman 2006; Furedi 2008). Citizens and organizations have been provided with a new risk language, causing them to evaluate different phenomena and activities in terms of risk. Thus, there is a dynamic relation between societal development and our understanding of and reflection on this development and society at large.

Within science we have witnessed a development from technical risk analysis – populated by philosophers, statisticians, and economists – to the broader field of risk governance, in which social scientists ponder how actors understand risks as well as how they handle them. Risks are put in specific contexts, which implies a call for social science in general and sociology in particular to develop knowledge on risk. This is the reason why we today can see an explosion of social scientific literature on risk; in particular how to analyze and manage risk.

The contribution of sociology to the field of risk research is mainly that society is differentiated, which means that also cognitions, understandings, and feelings of risk are differentiated. Actors have various cultural belongings and structural positions which make them understand reality differently, and therefore also act differently. Thus, to develop sociological knowledge on risks implies to contextualize risks; they are not the result of a calculation made beyond society, but instead the result of how actors, located in specific social settings, understand and manage certain phenomena. Risk is for sociology always a particular risk situated in a specific context.

Risk is a relative new object of sociological research, and even if this field has grown rapidly over the last three decades, it has not yet fully been institutionalized as a self-evident subfield of sociology (Krimsky and Golding 1992; Zinn 2008, p. 200). In addition, the sociological discipline covers a broad range of perspective and traditions, which is reflected in its research on risks; there are a number of sociological ways to conceptualize, understand, and conduct research on risk. Sociological thought spans from rational choice approaches to cultural theory; it encompasses micro-sociological theories on the construction of self-identities as well as macro-sociological theories on world systems. Thus, it is not an easy task to map out the sociology of risk. We will do so, however, by starting from some central assumptions of sociology, reviewing some of its most well-known approaches to risk, and finally by presenting a few important ongoing discussions and pointing out important areas in need of further research.

The aim of this chapter is to construe a sociology of risk that does not take technical risk analysis as a point of departure, to prevent sociology from being given a too restricted role in researching risk. Instead, we argue that the task of sociology is to contribute to an

understanding of the risk field in which risks always are situated in a social context and are necessarily connected to actors' activities. Thus, sociology opposes any kind of reification of risks, in which risks are lifted out of their social context and dealt with as something uninfluenced by the activities, technologies, and instruments that serve to map them.

The essay comprises four sections, this introduction being the first. The next and second section presents a historical perspective on sociology and risk. It starts by briefly describing what sociology is, followed by how the concept of risk gradually becomes an object of sociological thought. The section ends by presenting and discussing three different sociological perspectives on risk, all providing different contributions to the risk research field. The third section focuses on current strands of sociological risk research. It starts by giving an overview of three different sociological approaches to risk: Mary Douglas's cultural theory of social order, Ulrich Beck's theory of reflexive modernization and the risk society, and Niklas Luhmann's system theory. All three approaches conceptualize risk differently and make different contributions to the sociological study of risk. The review of these theories is followed by a presentation of five central, partly overlapping and ongoing discussions within the sociology of risk: risk governance, public trust, democracy and risk, the realism–constructivism debate, and governmentality and risk. Finally, the fourth section, based on these current discussions, briefly presents some areas in need of further sociological research.

## History

---

### What Is Sociology?

---

The origin of sociological thought can be traced to the end of the eighteenth century in Western Europe (Eriksson 1993). At this time, questions arose about the social order, the division of labor, social hierarchies, social cohesion, and individualization. A concept of society emerged that did not correspond with the sum of its population, but instead was a social phenomenon *sui generis*, a phenomenon with its own characteristics.

This understanding does not mean that society was external to human beings, but rather was something constitutive of them. The history of human beings and the history of society are two sides of the same coin. This perspective on society emerged with thinkers such as Adam Ferguson (1767), John Millar (1771), and Adam Smith (1776). Whereas Thomas Hobbes (1651) understood man as a “rational wolf” in need of external pressure from outside (in form of norms, laws, and force) to enable social life, these social thinkers understood human beings and society as interdependent. In their view, society was more than an external environment; it was also something inside us. Our words and deeds were not only individual acts; they were also social products. These thinkers were the predecessors of sociology, and 100 years later, the discipline of sociology first saw the light of day.

The growth of sociology is intertwined with the development of empirical data collection and social statistics (Calhoun et al. 2007, pp. 13–18). In a number of European countries, governments had begun to regularly collect information about their populations. Statistical analysis grew strongly, and the national census – originally a way to keep track of adult males’ availability for military service – emerged as a regular activity of the state, taking its modern form in the nineteenth century. British Parliamentary investigations of industrial conditions provided the empirical basis for Karl Marx’s initial theorizing, empirical

data on deaths collected by governments and churches for Émile Durkheim's study of suicide, and publicly gathered data for Max Weber's investigation of German peasants and "junker capitalism."

Sociology was constituted as an empirically based social science, with an emphasis on the importance of context (social, material, economic, cultural) for understanding social life. It also emphasized that society was a social phenomenon in its own right; it was not just an aggregate of individuals. Social practices and collective understandings were not possible to explain by referring to individual human beings.

Phenomena and activities should be understood and explained in relation to their social contexts. And this applies not only to norms and artifacts, but also to actors and knowledge. Hence, sociology is critical of individualistic explanations, though without rejecting the importance of actors. Individual human beings are always situated in social settings, which have to be considered when explaining their cognitions, feelings, and practices.

## Risk in Sociology: From Social Problems to Risk

---

The relationship between individuals and society has been central to sociological thought since its origin (Giddens 1984). There have been – and still are – many ways for sociology to explore and explain this relationship. Even if everyone agreed that there was no such thing as "pure individuals" – human beings unaffected by society whose thoughts, emotions, and wills developed apart from society – they all emphasized that this relationship was not a harmonious one. People find themselves limited by social positions and cultural belongings and struggle to transform structural barriers and cultural restrictions. At the same time, sociologists emphasized that human beings' aspirations and ideals come not only from inside, but also from outside – from social norms and ideals that surround them and which they gradually have internalized. They do not develop their own goals, values, and preferences apart from those that exist in society, but instead develop them in relation to these. Thus, the human being is neither a puppet, nor her own master. Social structures and cultural belongings not only serve as barriers to social action, they also enable it.

In classical sociology, social problems and not risks were the focal point. Its classical thinkers – Karl Marx, Max Weber, Émile Durkheim, and Georg Simmel – were preoccupied with the emergence of modern society, not least the development of industrialization, urbanization, and rationalization and their degrading effects on human beings. Different kind of social problems were put to the fore in the sociological analysis, and different angles were tried in exploring these problems. Risk was not incorporated as a conceptual lens through which these problems were understood and analyzed. Instead, risk research emerged and developed without any relation to sociological thought. The reasons for this were dual: disciplines dealing with risks did not see any relevance of sociological analysis, and sociology did not see risk as a relevant object for sociological research.

Traditional risk concepts have been developed within a framework where risk is technically defined. For *technical risk analysis*, risk means to anticipate potential harm to human beings, cultural artifacts and ecosystems, to average these events over time and space, and to use relative frequencies (observed or modeled) as a means to specify probabilities (Renn 1998, p. 53). Thus, risk concerns a situation or event in which something that human beings value is at stake and where the outcome is uncertain (Jaeger et al. 2001, p. 17).

This kind of analysis implies that one set of experts establishes the probability and magnitude of the hazards and another set of experts evaluates the costs and benefits of various options. Thereafter, political priorities are invoked in order to make decisions on regulating (forbidding, controlling, permitting) certain risks (Amendola 2001). Science is pivotal in measuring and assessing risks, and therefore experts should guide the risk management processes.

Thus, technical risk analysis is an un-sociological understanding of risk; it does not consider the broader social, cultural, and historical context from which risk as a concept derives its meaning (Lupton 1999, p. 1). In response to this kind of analysis, three different sociological perspectives have emerged, all making different sociological contributions to the field of risk research: the social construction of misperception of risk, the social amplification of risk, and the social construction of risk. With this development, important ideas from classical sociology have gradually been taken advantage of and made to influence risk research.

## Sociology Explaining Public Misperceptions of Risk

---

Risk researchers and risk managers gradually recognized that the public's perception of risk was different from the view held by the experts. The nuclear researcher Chauncey Starr's seminal article "Social benefit versus technological risk," published in *Science* in 1969, emphasized the importance of considering public acceptability of risk (Starr 1969). He found that risk tolerance was correlated with a number of social components. For example, the public more easily accepted voluntary and familiar risks than involuntary risks (comparing risks associated with the same level of social benefit).

Social and behavioral scientists have devoted themselves to finding out how different groups and individuals perceive risks (Gutteling and Wiegman 1996; Breakwell 2007). Their point of departure – most explicitly within the *psychometric school of risk analysis* – is that for laypersons risk is a subjective assessment in which contextual factors play an important role. This does not mean that citizens' reasoning is irrational or haphazard. On the contrary, it is possible to find cognitive patterns and trace causal factors that explain citizens' risk perceptions. It is found that factors such as novelty (how new a risk is), dread (how feared the risk is), and if the cause of the risk is seen as tampering with nature, are significant factors shaping risk perception (Slovic 1987; Sjöberg 2000). Perceived influence and power are also important factors, as is cultural belonging (Finucane et al. 2000; Zinn and Taylor-Gooby 2006). Also of importance is the social and spatial context in which people make judgments (Lidskog 1996; Wester-Herber 2004). Much research has also been concerned with how different social groups access, interpret, understand, and respond to different forms of information in diverse contexts (Slovic and Peters 1998; Bickerstaff and Walker 2001; Howel et al. 2002).

Thus, there are a number of contextual and social factors that explain why the public does not assess risk in a similar way as the experts. This perspective considers citizens' perception and understanding of risk, and does not give any attention to the perception of experts and how they understand and measure risks. Instead, the technically defined risk is taken for granted; it is seen as an objective phenomenon in which scientific measurements and statistical calculations give correct, or at least the most valid, knowledge on the character of the risk.

When technically defined risks are not seen as contextually generated, the public's risk understanding is portrayed as biased or incorrect compared to the experts' more accurate

assessment. The difference between expert and citizen understanding is interpreted as caused by public ignorance or misunderstanding of science (Irwin and Wynne 1996; Levinson and Thomas 1997). By informing – and sometimes even educating – laypeople about the “real risk,” it is believed that the public would correct its judgment and accept risks that experts and regulators have found to be acceptable (Gouldson et al. 2007). According to this view – commonly labeled as *the deficit model of public understanding of science* (Irwin and Wynne 1996) – knowledge is first produced in a closed circle of scientists, after which it should be disseminated to the public (who in many cases are unable to understand science properly). This view is a variant of the “sociology of error,” which explains what is seen as error and falsehood in science with reference to contextual factors (such as traditions, ideology, conventions, authority, and interests) and what is seen as true and valid knowledge with reference to observations and reasons. The role of sociology, however, is to explain error, not truth (Bloor 1976).

An implication of this perspective is that risk assessment concerns objective analysis aiming to produce factual knowledge about specific risks, while risk communication is about the distribution/transmission of this factual knowledge to the public. To make risk communication effective, it is important to understand how different segments of the public understand risks, and assess the sources of information, to be able to effectively inform them about risks in different circumstances. The sociological task is to provide knowledge about these factors that result in public misperception of risks.

## Sociology Explaining Amplifications of Risk

---

The deficit model draws a sharp line between risk as defined and assessed by experts and as understood by the public. However, many sociologists seek a more sophisticated way to understand the clash between experts’ and the public’s understanding of risk. In the late 1980s, the *Social amplification of risk* approach was developed (Kasperson et al. 1988; Pidgeon et al. 2003). This is a communication model according to which an original physical event generates a signal that passes different social stations that amplify or attenuate the signal. The model explains why risks evaluated by technical risk analysis as being similar may receive different levels of attention in society at large.

The basic assumption is that a risk event has certain physical characteristics, such as material damage, injuries, and deaths. These characteristics provide an original signal that is then transformed in the communication process. The risk event itself has no meaning, but the social stations of amplification charge it with meanings and messages. The social amplification of risk explains how risks and risk events interact with psychological, social, institutional, and cultural processes in ways that amplify or attenuate risk perceptions and public concern, and thereby shape risk behavior.

It also explains the development of secondary effects of risk events, that is, risks caused by the amplification processes and not by the signal itself (Kasperson et al. 1988). Social amplification of a risk event associates it with meaning that may result in changed policy regulation, new conditions for insurance, consumer boycotts of a product, decreased institutional confidence, and social stigma. Thus, the amplification may result in social or economic consequences that go far beyond the direct consequences of the risk event.

Thus, technical risk analysis cannot provide information about how risks are amplified in society, because understanding and explaining processes of amplification is solely a task for

social and behavioral science. The social amplification approach proposes a division of labor in which technical risk analysis is concerned with investigating the original signal whereas social science in general and sociology in particular analyze how this signal is transformed by society.

The social amplification of risk approach is symmetric in the sense that both the attenuation and intensification of the original risk signal are taken into account. Both technical risk analysis and sociological analysis are needed in order to gain knowledge on risk and its consequences for society. The approach contributes an understanding of why certain hazards and events that experts assess as low risk may receive public attention, whereas other hazards that experts consider more severe receive less attention (Kaspelson et al. 2003). By encompassing different factors on different levels, it presents a dynamic and multilayered view on how risk understanding develops. It also aims to link the three leading schools of risk analysis – technical risk analysis, psychometric studies of risk perceptions, and sociocultural studies on risk understandings – into a single framework (Pidgeon et al. 2003).

This bridging effort is ultimately based upon a clear division between a physical world of events and a social world of meanings. The amplification process starts with a physical signal, either in the form of an event (e.g., an earthquake) or the recognition of an adverse effect (such as the discovery of climate change). Thereafter, social factors attribute meaning to it. Risk is thereby conceptualized partly as an objective property of a hazard or risk event, and partly as a social construct (Kaspelson 1992, p. 158). According to its proponents, this position avoids the two problems of conceptualizing risk in a totally objectivistic or in a relativistic manner (Renn 2008, p. 39).

The approach links different ways to understand and analyze risk. It is, however, a synthesis based on a linear model in which something external to society is channeled through amplifying stations, resulting in different consequences, and where feedback mechanisms and processes of iteration only take place between the social stations of amplification. The hazard itself and experts' calculation of risk (not least through technical risk analysis) are left outside the approach and are not included in the analysis. In this way, risks – or at least risk events and hazards – are positioned as external to society. Fact finding and sense making are seen as different and discrete spheres of activity, the former populated by technical risk analysts and the latter by various segments of the public. This model does not discuss how the risk (in form of risk events, hazards, or the technical calculation of risk) is constructed, but only how it is amplified. Hence, the bridging ambition also results in a reproduction of the divide between expert and public understandings of risk. Not only risks are amplified in this approach, but also the divide between risk and understandings of risk.

## Sociology Explaining Risk

In contrast to sociological studies of the misperception of risk and the amplification of risks, the *social construction of risk* approach includes the role of science and technical risk analysis as topics to investigate. Its starting point is that all risks are socially constructed in the sense that risks always exist in contexts (Wynne 1992a). This means that technical risk analysis and experts' assessments of risks have no privileged position; they are only one of many possible ways to frame, define, and understand risks. Thus, knowledge is intimately related to meaning and actors, which means that no kind of knowledge and no kind of actors should be excluded from sociological analysis. Instead, a symmetrical approach is put forward, where all

risks – irrespective of how they are assessed and by whom – are seen as socially constructed. Risks, hazards, and risk events are all sociocultural phenomena in their own right, and should not be seen as unproblematic facts that generate specific signals which laypeople then misunderstand or social stations amplify.

This understanding – that also science's assessment of risk should be seen as construction of risk – has been fuelled by recent developments in society. Science was initially applied to a “given” world of nature, people and society, and scientific skepticism demystified the social and natural worlds. Science's own claim of rationality was itself spared from the application of scientific skepticism. According to Ulrich Beck (1992), a process of “reflexive scientization” is gradually taking place, whereby scientific skepticism is extended to consider the inherent foundations and external consequences of science itself. This demystification opens up new possibilities for questioning science and technical risk analysis. This extension of the scope of rational skepticism means that no scientific statement is “true” in the old sense of there being an unquestionable, eternal truth, where “to know” means to be certain (cf. also Giddens 1990, p. 40; 1994).

The implications of this perspective – risk as a product of social processes – are far reaching. Not only does it mean that the task of sociology is to analyze how actors – including science and risk experts – frame, define, understand, and manage risks. It also implies that the separation of risk regulation into distinct areas – risk assessment, risk management, risk communication – is incorrect. Values are not solely invoked in the initial process of defining risks that then should be analyzed, evaluated, and regulated by technical risk analysis. Instead, they are an intrinsic part of the risk regulation process, as in the process of developing and validating knowledge. Thus, even if they are presented as separate spheres, they are not discrete activities ordered in a linear process aiming to regulate risk, but instead are dynamically related to each other. The problem is that technical risk analysis' definition of risk is preceded by an implicit framing, which is rarely the subject of discussion, either by citizens or by the risk researchers themselves. This framing provides a very restricted understanding of risks and of actors, behaviors, and processes (Wynne 1992a, 2005). When this framing is naturalized – taken as a pre-given way to understand and conceptualize risk – it restricts the role of sociology to investigate and explains why actors' perceptions and understandings of risk differ from those put forward by technical risk analysts.

The social construction of risk approach has been criticized, not only by technical risk analysts, but also by social scientists. To consider risk as a social construct implies, according to its opponents, a far-reaching relativism where risk bears no relation to a reality beyond human consciousness and cultural values. The social amplification of risk was explicitly developed with the aim of transcending the division between naïve empiricism and far-reaching relativism, as an approach that includes both the need for technical risk analysis and the need for cultural theory (Kasperson et al. 1988). Some researchers, such as Ortwin Renn (2008), argue that it is possible to take advantage of both a more contextual understanding of risk and a more traditional analytical and context-less approach to risk. This is done by defining risk as constituted by both physical/material and social/cultural elements (Kasperson 1992, p. 158; Renn 2008, p. 2). This argumentation rests on the general assumption that it is possible to analytically separate values and evidence, social norms and factual knowledge, deliberation and analysis, but that in practice there is a need for better integration of these analytically distinct entities.

As will be shown below, to understand and analyze risk as a social construct does not necessarily imply a strong relativism, but is based on the assertion that risks are social facts that

are irreducible to technical measures. Similarly, its view of knowledge – as contextual, unstable, and sociocultural – does not imply a relativism where all standpoints are given the same cognitive value.

## Current Research

---

A number of sociologists, from somewhat differing standpoints, have emphasised that risk has largely replaced the previous notions of fortune and fate (Beck 1992; Bauman 1993; Luhmann 1993; Giddens 1999). In the past, a lack of certainty was attributed to powers (God, nature, magic) beyond human control, whereas today it is attributed to organizations (such as scientific communities, companies, and nation states). Risk is a factor in human decision making because we cannot gain sufficient knowledge about which possible future will result from our decisions. Furthermore, risk is constituted by the distinction between present reality and future possibilities. Thus, it presupposes that the future is not determined, and that human action shapes the future. As Anthony Giddens (1990, p. 3) puts it:

- ▶ Modernity is a risk culture... The concept becomes fundamental to the way both lay actors and technical specialists organise the social world. Under conditions of modernity, the future is continually drawn into the present by means of the reflexive organisation of knowledge environments.

However, to reflect on future consequences of human action is nothing new in history. The decisive difference is that in modern societies almost all aspects of social life are included in these reflections and have become objects of decisions and deliberations; hence thinking in terms of risk assessments is a more or less ubiquitous exercise in everyday life (Callon et al. 2009).

In the following, we will present three important sociological contributions to risk research, all of which treat risk as central for society at large. They take into account the broader historical, social, and cultural contexts from within which risk derives its meaning and resonance. Thereafter, we present five thematic areas that are objects of lively discussion in contemporary risk sociology.

## Important Theories

---

### Mary Douglas: Purity and Danger

The British anthropologist Mary Douglas (1921–2007), who has inspired many social scientists in the field of risk research, argues in her book *Purity and Danger* that risks should be understood with reference to the social organization (Douglas 1966). The assessments of risks are responses to problems in the social organization of a specific society, but are also resources for building social order and defending social boundaries. The way risks are viewed reflects the organization of society, including its borders to other societies. What we usually understand as threats coming from outside of society are in fact problems within society.

More specifically, Douglas is interested in how societies assess purity and pollution, which she connects with the overarching concept pair of order and disorder. Purity supports order

(both cognitive and social) and pollution is what deviates from and threatens order, and should therefore be condemned. According to Douglas, the separation between purity and pollution, the latter signifying danger, is one of the most fundamental conceptual distinctions in our thinking. This division is, however, relative: "There is no such thing as absolute dirt: it exists in the eye of the beholder" (Douglas 1966, p. 2).

It is important to stress that these definitions are not chosen individually, but in a collective process which is compelling for individuals. Demands for purity are simultaneously requirements for social order, for the survival of society. Douglas argues that we all have norms of order and an associated type of purity to defend, but that these orders and types vary between groups and societies.

In every defense against risks, there is a wish to protect a social order that is considered endangered. Discussions about risks include a desirable norm – a norm of purity – from which the seriousness of the risks can be established. This norm makes it possible to require risk reduction and thereby increases purity. Demands for better risk management imply demands for societal change. A norm of purity contains a vision of a societal order that better corresponds to this norm. Therefore, for sociology, risk should never be seen as something out there, separate from society, but as something produced in and by society.

One implication of this perspective is that our use of the concept of risk reveals who we are. Values and beliefs (including preferences and knowledge) – what Douglas calls *cosmologies* – are viewed as coherent and endogenously derived systems (cultural biases), generated from specific patterns of social relations (Douglas 1978). Such cosmologies support and legitimate social relations. Actions, organizations, and knowledge interact to generate and legitimate social relations. This interplay should not be understood as a unidirectional causal relation but as a reciprocal interaction in which knowledge and society are co-produced (cf. Jasenoff 2004).

One well-known tool for interpreting and explaining such co-production is the *grid–group typology* originally developed by Mary Douglas and her coworkers (Douglas 1982, 1996; cf. Thompson et al. 1990). "Grid" describes the internal structure, how roles and activities are positioned, and "group" the external borders, how the boundary between insiders and outsiders is defined. These two dimensions are fundamental to all cultures, and imply four different cultures: hierarchical, individualistic, egalitarian, and fatalistic.

*The hierarchical culture* is characterized by a stable and regulated internal social order (high grid). Group membership is strong (high group); it is clear to everyone who is a member and who is not. This culture is characterized by formal procedures, rules, routines, timetables, and trust in authorities. *The individualistic culture* is the opposite of the hierarchical. Group membership as well as hierarchy are low (low grid and low group). This is a culture of enterprise characterized by uncertainty and change. Decision making is performed with a minimum of formal procedures, and is based on trust in individual competence. *The egalitarian culture* has a strong boundary to other groups and the outside world (high group). Purity is strived for and outsiders are viewed as threats. The internal differentiation is low (low grid); equality is strived for between positions. *The fatalistic culture* is a residual culture. It is neither individualistic nor collectivistic, but rather cut off (low group). It includes individuals who do what they are told, though without the protection of either social privileges or individual skills (high grid). Those who govern activities and formulate plans for what will happen are always someone else.

Douglas has used this typology to explain the existence and distribution of different risk perceptions (Douglas and Wildavsky 1982; Douglas 1992). The way individuals and groups

react to risks reveals their cultural belongings. Is the reaction about *embracing*, *ignoring*, *rejecting*, or *adapting* (Douglas 1978)? In an uncertain situation of risk, are the possibilities emphasized or the negative consequences? These questions are relevant for all kinds of issues that people see as risks and dangers; they can concern things like nuclear power and biotechnology, but also EU membership and immigration.

In terms of the grid–group typology, Individualists tend to focus on the possibilities, *embracing* risks as opportunities to exploit for personal profit. Fatalists do not know how to react and tend to *ignore* the risks. Egalitarians mobilize resistance in order to *reject* and eliminate the risks. Hierarchical people try to assimilate and *adapt* to the risks through regulation and control of risk activities.

## **Ulrich Beck: The Risk Society and Reflexive Modernization**

Ulrich Beck's (1992) *Risk Society: Towards a New Modernity* – originally published in German in 1986 – is one of the most influential works of social analysis in recent decades. This book is about the reflexive modernization of industrial society. Beck's underlying thesis is that we are not witnessing the end but the beginning of modernity, a modernity beyond its classical industrial design. This guiding idea is developed from two angles. First Beck focuses on the social transformation from an industrial society with its production of wealth to a risk society with its production of risks and social hazards. The other side comes into view when Beck places the immanent contradictions between modernity and postmodernity within the industrial society at the center of discussion. Thus, risk and reflexive modernization are the two – intrinsically interrelated – themes of this book. Beck explicitly states that the aim is to seek “to understand and conceptualize in sociologically inspired and informed thought these insecurities of the contemporary spirit, which it would be both ideologically cynical to deny and dangerous to yield to uncritically” (Beck 1992, p. 10).

Just as modernization in the nineteenth century dissolved the structure of feudal society and produced the industrial society, modernization today is dissolving industrial society and another society is coming into being. This new and coming society is “the risk society,” which is a distinct social formation just as the industrial society was. The risk society differs very clearly from the industrial class society in that it focuses on the environmental question and the distribution of risks instead of the social question and the distribution of wealth. In both types of societies, risks are socialized, that is perceived as a product of political decisions and human action; but in contrast to the risk society, the classical industrial society saw risks as manageable side effects of the production of wealth. These risks were legitimated partly with reference to the production of wealth and partly through society’s development of precautions and compensation systems.

Risk is defined by Beck (1992, p. 21) as “a systematic way of dealing with hazards and insecurities induced and introduced by modernization itself”. The risks and hazards of the risk society are different than in the industrialized society, as they are more widespread and serious. For the first time in history, society involves the political potential for global catastrophes. In the risk society, the relation between wealth production and risk production is reversed. The production of wealth is now overshadowed by the production of risks. The risks produced have lost their delimitations in time and space and consequently can no longer be seen as “latent side effects” afflicting limited localities or groups.

At one level, the distribution of risks adheres to the class pattern, but it does so inversely: wealth accumulates at the top, while risks accumulate at the bottom. Therefore, the risk society could be seen as simply strengthening the class society. However, on another level, this is not true. Today's diffusion and globalization of risks entails "an end of the other"; that is private escape routes shrink (it is impossible to buy yourself free from risks) as do the possibilities for compensation. Thus, risk positions are no longer pure reflection of class positions, but instead they transform and replace class positions. One example of this is that property (such as forests) today is being devaluated; it is undergoing a creeping "ecological expropriation" which implies the emergence of new conflicts between the different interests of profit and property. This means that the central conflicts in the future will be not between East and West, between communism and capitalism, but between countries, regions, and groups involved in primary and in reflexive modernization, the latter being those that are striving to relativize and reform the project of modernity.

Global risks – mega-hazards, to use Beck's term – overlap with social, biographical and cultural risks, as well as insecurities. Today, these latter forms of risk have reshaped the inner social structure of industrial society and its fundamental certainties of life: social classes, familial forms, gender status, marriage, parenthood, and occupations. This comprises the other part of Beck's discussion of reflexive modernization.

The theory of modernization is formulated by Beck as the unleashed process of modernization overrunning and overcoming its own "coordinate system". This coordinate system has fixed the understanding of the separation of nature and society, the understanding of science and technology, and the cultural reality of social class. It features a stable mapping of the axes between which the life of its people is suspended – family and occupation. It assumes a certain distribution and separation of democratically legitimated politics on the one hand, and the "subpolitics" of business, science, and technology on the other.

Today, a social transformation is underway within modernity, in the course of which people will be set free from the social forms of industrial society. This reflexive modernization dissolves the traditional parameters of the industrial society (such as class and gender). This "detraditionalization" occurs in a social surge of individualization, through which a capitalism with individualized social inequality is developing. Here the family is replaced by the individual as the reproductive unit of the social in the life world.

Having discussed this theory of individualization, Beck turns to the role of science and politics in the era of reflexive modernization. He argues that when encountering the conditions of a highly developed democracy and well-established scientification, reflexive modernization leads to an unbinding of science and politics. Earlier monopolies of knowledge and political action are then differentiated.

In discussing science, Beck makes a distinction between primary and reflexive scientification, with the former meaning that science is applied to a "given" world of nature, people and society, and the latter that the scope of scientific skepticism is extended to encompass the inherent foundations and external consequences of science itself. Reflexive scientification thus entails a demystification and demonopolization of scientific knowledge claims. At the same time, the role of science in the risk society is growing. Today's threats are beyond human perception and experience and it is through science that risks become known. Science thus comprises the "sensory organs" for the perception of today's risks. Taken together, these two parallel and different developments do not mean that science has come to an end; on the contrary, it pervades all areas of modern life. Today's science is undergoing a situation of being

dethroned similar to that which happened to (institutionalized) religion. In Beck's secularization model of modern science, the future will bring about a pluralization and a marketization of science.

Beck has been of pivotal importance for sociology, not least by paving the way to make risk a central concern for general sociology. In the wake of *Risk Society*, he has published a number of books in which he further develops his perspective: *Ecological Politics in an Age of Risk* (1995), *Ecological Enlightenment* (1995), and *World Risk Society* (1999). In his later writings – such as *What Is Globalization?* (2000), *Individualization* (2002, with Elisabeth Beck-Gernsheim), *Cosmopolitan Vision* (2006), *Power in the Global Age* (2006), *The Brave New World of Work* (2010), and *A God of One's Own* (2010) – Beck puts more emphasis on reflexive modernization and its importance for all aspects of society such as family, work, religion, and global politics. However, in all his books, irrespective of their subject, the main theme is reflexive modernization and the future development of society.

### Niklas Luhmann: System Theory and Risk

Drawing on Talcott Parsons's social theory, the German sociologist Niklas Luhmann (1927–98) developed a general theory of modern society. The starting point is that there is a fundamental distinction between system and environment and communication is the basic social operation (Luhmann 1984, p. 47). A higher complexity in the environment entails a greater importance for the system to reduce this complexity, otherwise the system will not be operational.

This reduction of complexity is accomplished through functional differentiation, which means that different subsystems develop, each with distinct forms of communication (programs and binary codes). These subsystems are self-referential and autopoetic, which means that their internal orders guide their observations and interpretations and that they are not formed and structured by any external factors.

A subsystem is cognitively open; it is receptive to signals from its environment. At the same time, it is operationally closed. Signals are always transformed into communication through a particular binary code of the subsystem. These codes are abstract and universally applicable distinctions. Science codes a signal in terms of truth/untruth; economy in terms of property/no property; law in terms of legal-illegal; religion in terms of transcendent/immanent; and politics in terms of political power or lack of power and so forth. Luhmann (1989, p. 18) states that:

- ▶ The system introduces *its own distinctions* and, with their help, grasps the states and events that appear to it as *information*. Information is thus a purely system-internal quality. There is no transference of information from the environment into the system. The environment remains what it is.

This does not mean that nothing else exists than social systems and their communicative processes. What it says is that external facts can only be taken into account as part of the system's environment and only be understood through communication. There is no position available outside the system; these facts can only be understood from within (Luhmann 1993, p. 5). It is, however, possible to observe the border of a system, and this is done through "second-order observation" (Luhmann 1993, p. 223). First-order observations identify facts and objects as givens, and do not reflect on the distinction used in the observation. Second-order observation is an observation of the distinction implicitly used in the first-order

observation; it recognizes which distinction is applied in observing a fact or object. Luhmann (1993, p. 227) stresses that second-order observation is not more true or objective than first-order observation. What second-order observation reveals is that there are no objective facts outside the operation of each subsystem. What may seem like an objective fact in the first-order observation (which takes for granted its own distinction), is a product of a particular distinction made in the observation process. Thus, the same event is coded differently by different subsystems.

For instance, the signal from the Tohoku earthquake in Japan, March 11, 2011, that resulted in more than 15,000 deaths and a nuclear disaster at Fukushima nuclear power plant is coded radically differently by different subsystems. The economic subsystem focuses on price mechanisms, economic compensation, and falling prices of shares; the political subsystem on political legitimacy of decisions concerning the location of the nuclear power plants and how the disasters were handled by authorities, but also on the legitimacy of the political representatives that had permitted this activity; the legal subsystems on violations against the given permissions for the plant and the liability of the company as well as political institutions; science on health consequences of radiation exposure for workers and the local population. It is what is communicated that counts. A phenomenon, an event, or an activity can never in itself create a response; it needs to be subject of communication.

- ▶ But as physical, chemical, or biological facts, they create no social resonance as long as they are not the subject of communication. Fish or humans may die because swimming in the seas and rivers has become unhealthy. The oil pumps may run dry and the average climatic temperature may rise or fall. As long as this is not the subject of communication, it has no social effect. Society is an environmentally sensitive (open) but operatively closed system. Its sole mode of observation is communication. It is limited to communicating meaningfully and regulating this communication through communication (Luhmann 1989, pp. 28–29).

Risk is inherently linked to a functionally differentiated society. In contrast to many other theories, Luhmann does not see risk as a result of detrimental activities or as caused by industrial society. Instead, risk is attributed to decision making that may result in negative consequences. Contingency is a central concept for Luhmann, which means that a situation includes a large number of possibilities. To be able to act, it is necessary to choose among these possibilities, but there is no fundamental point – external authority – that tells what to select among the various alternatives. This has to do with the development of the functionally differentiated society.

In contrast to earlier societies, there is no privileged function system in society, which means that a functionally differentiated society has no center. Each subsystem can only refer to its own communication, and only internally can it refer to its environment. Through internal differentiation, the system develops a richer way to manage the complexity of the environment. At the same time, this differentiation results in a higher internal complexity, with different functional subsystems existing side by side and communicating with their own specific codes and with no external authority. Earlier societies' external references (such as religion) are replaced by the social system's self-references in the form of subsystems.

Luhmann defines *risk* as an attribution of an undesired event or possible future loss. Thus, risk is an intrinsic part of a functionally differentiated society. Decisions have to be made without any certainty about what consequences they will lead to. The cause of the damage could either be attributed to the system itself (risk) or something external to the system

(danger) (Luhmann 1993, pp. 101–102). This means that risks concern attribution, which becomes even more clear when Luhmann discusses another distinction, namely, between those who take the decision and those who are exposed to its consequences (Luhmann 1993, pp. 105). Those who take the decision face a risk; whereas those who are victims face a danger, that is, those who perceive themselves as exposed to something that they cannot control. Uncertainty is intrinsic to both risk and danger; the difference lies in who is seen to be a decision maker.

Luhmann (1993, p. 109) stresses that “one man’s risk is another man’s danger” and claims that there is a growing gap between those who participate in decision making and those who are excluded from decision-making processes but have to bear the consequences of these decisions.

Luhmann’s system theory provides an alternative understanding, not only of risk but also of society at large. It sees risk as a matter of attribution and communication, associates it with decision making, sees it as inherently linked with a functionally differentiated society, and strongly emphasizes the distinction between risk and dangers and between decision makers and those affected; and in doing so the theory has both received support and met with criticism (Japp and Kusche 2008, pp. 101–103).

## Thematic Areas

---

There is today an ongoing and lively discussion within the sociology of risk. In what follows, we present five partly overlapping areas of central importance in contemporary risk sociology: risk governance, public trust, democracy and risk, the realism–constructivism debate, and governmentality and risk.

### Organizational Risk: From Risk Analysis to Risk Governance

To conceptualize an object as a risk entails seeing it as manageable and governable (Baldwin and Cave 1999; Hood et al. 2001; Hutter 2001; Lidskog et al. 2009). Risk creates space for action as it opens the future for calculation, deliberation, and decision making. In this sense, regulation “enrolls” futures and shapes policy formulations (Wynne 1996).

Risk regulation is not only about how to govern an existing reality, it also concerns the transformation of this reality, for instance, by dealing with novel forms of knowledge that have not yet been put into industrial practice (cf. Stehr 2005). Regulation does not only concern how to regulate an existing activity, but also how novel knowledge should be deployed and employed.

As shown above, there has been far-reaching criticism of technical definitions of risk, not least concerning the difficulty of upholding a sharp separation between an objective measure of risk and a sociocultural understanding of risk (Hilgartner 1992; Rosa 1998; Amendola 2001; Todt 2003). The critique was initially directed at problems *within* risk analysis, and consequently public perception of risk was seen as a challenge to how risk was defined and approached by technical and calculative means. In the 1990s, however, there was a shift from this internal focus to the broader question of the legitimacy of government. Organizations must ponder not only how to deal with technically defined risks, but also how to deal with

actors who may question both the legitimacy of current methods for regulating risk and the trustworthiness of organizations responsible for this regulation. Prompted by several regulatory failures, authorities and companies have started to account for and deal with public opinion and public perceptions of risks, not only to handle criticism but also to forestall it (Löfstedt 2005). Programs for risk communication, public relations, stakeholder dialogue, and public involvement are today integral to both public and corporative governance (Gouldson et al. 2007). Calls for more inclusive and transparent processes, public dialogue, and democratic engagement are widespread in society (Irwin 2006). A heightened concern for stakeholder involvement and public inclusion can be seen as a strategy to influence perception, shape understandings, and produce legitimacy.

Michael Power describes this shift as a move from risk analysis to risk governance. The “governing gaze” has shifted from how risk is defined, analyzed, and calculated to the governance of the organizations that analyze risk (Power 2007, p. 19). Even though the call for a more inclusive risk analysis and risk management may have evoked some response from regulatory agencies, it has not directly led to more inclusive and deliberative risk regulation processes. Rather, the shift from risk analysis to risk governance has increased the awareness of how organizations deal with public opinion and public perceptions as a source of risk in the sense that such perceptions could pose a threat to the legitimacy and stability of existing ways of governing risk (Power 2007, p. 21). This justifies research on how organizations deal not only with technically defined risks but also with the actors they perceive as possible threats and potential risks to the stability of the organization.

Risk regulation does not only concern what is acceptable in terms of how we should mitigate or accept certain risks and hazards, but also rules regarding the process itself and activities that target the understanding of risk and deal with public opinion and perceptions concerning it. Thus, risk governance is not limited to technical calculation of risk, but also includes the evaluation of organizational aspects in regulation of risk.

With a focus on risk governance, that is how uncertainties are organized in order to transform them into governable risk, the questions of who should be involved or excluded in risk regulation processes, on what grounds, and what aspects should be made open and transparent to others, gain greater relevance due to the legitimacy gains and losses such decisions may generate. A heightened concern for public involvement in regulation can thus be seen as “a strategy to govern unruly perceptions and to maintain the production of legitimacy in the face of these perceptions” (Power 2007, p. 21).

## **Public Trust: The Relation Between Experts and Laypeople**

Technical risk analysis builds on a sharp boundary between experts and laypeople. Laypeople do not have access to all the knowledge possessed by experts and therefore draw different conclusions about risks, their ordinariness, magnitudes, and impact. In technical risk analysis, scientific knowledge is the norm and this is what experts have but laypeople lack. This difference is what motivates the concept of lay knowledge, of not being an expert. Focusing on what laypeople lack constitutes the basis for *the deficit model* mentioned earlier in this essay (Irwin and Wynne 1996). According to this model, the solution is to inform and educate laypeople in order to give them the capacity to gain correct knowledge and thereby arrive at the same conclusions as experts. Risk psychology and risk communication originally developed as

academic fields with the aim to understand how laypeople reason about and assess risks in order to learn how to effectively communicate correct knowledge to this group.

The deficit model has been heavily criticized, not least by researchers in the field of science and technology studies (STS) (Wynne 1995; Irwin and Wynne 1996). These scholars argue that the most important problem is not that the public is unaware of research results and scientific facts, but that scientific experts are unaware of and disinterested in lay knowledge and how laypeople assess the situation when decisions are to be taken on complicated risk issues. Consequently, the problem is not that laypeople lack knowledge or lack trust in expertise, but that experts in technical risk analysis do not trust laypeople. Laypeople have the competence to contribute to discussions and decisions on risks, since these concern much more than scientific facts. If grasping scientific details becomes the most important requirement for participation in risk discussions, the relevance of scientific knowledge becomes heavily exaggerated (Irwin and Michael 2003, pp. 22–28). Despite their lack of scientific knowledge, laypeople are competent actors with developed abilities to reflect on what types and sources of knowledge are of relevance to both risk analysis and risk assessment and why some experts should be more trusted than others.

Today, public involvement is often devoted much attention, but there is a tendency to frame this involvement from a technocratic understanding based on the deficit model (Irwin 2006; Lidskog 2008). In this way, broadened participation gives experts further possibilities to inform the public with the aim of winning acceptance for already proposed decisions. This instrumental ambition can be found in every participatory project, because there are always groups who strive for a specific outcome of the process. Studies have found that when laypeople are not considered competent to influence the decisions – when they are taught instead of listened to – the result is often one of alienation rather than engagement amongst the public (Wynne 2001).

Problems arise when such an overconfident and self-sufficient expert culture tries to communicate the benefits of a risk project. This culture is not interested in reflecting on its own shortcomings, and criticisms from laypeople are understood as based on their not understanding what is best for them (Wynne 2001, p. 447). If expert cultures wish to increase their legitimacy and appear as trustworthy to the public, they should instead be less confident about their own results and open to acknowledging their own limitations.

It is, however, not only uncertainties to which attention should be given attention, but also ignorance. Scientific knowledge is strongly specialized with a narrow focus, which implies that complexities are reduced and alternatives are actively deleted. In order to increase the trustworthiness of scientific knowledge, it is important to make visible the conditions of scientific knowledge production and risk assessment. The implication here is that scientific knowledge alone is not enough when deciding about complicated risk issues. It must therefore be enriched by other types of knowledge, as well as by other perspectives in order to give a more complete and nuanced view of the risks at stake.

The alternative to the deficit model and a technocratic framing of risks is to include a broader understanding of participatory processes, one which acknowledges that other actors than scientific ones can contribute knowledge on risk and therefore should be given possibilities to influence the regulation of risks (Funtowicz and Ravetz 1990; Lidskog and Sundqvist 2011). Hence, questions concerning who formulates the issue, sets the agenda, and exercises power become important. However, this does not mean to replace blind trust in experts with blind trust in laypeople. To draw clear boundaries between experts and lay people and grant

one of these priority over the other is not the right way to proceed. Instead it means that the public *always* has important contributions to make in technical discussions on risk issues. These contributions should never be evaluated from the deficit model for the reason that they are not about scientific facts, but about how to assess the relevance, trustworthiness, and ignorance of scientific facts. This kind of public competence, which emphasizes the contextual dimension of science, can always enrich scientific knowledge (Wynne 1993, p. 328).

The public can also contribute knowledge and insights about what they are worried about (Marres 2007; Sundqvist and Elam 2010; Lidskog 2011). This knowledge as well as the assessments made by members of the public are anchored in their livelihoods. Experts and authorities are often completely unaware of the reasons for ordinary people's worries, and it is therefore of great importance to involve concerned groups in decision processes in order to include relevant experiences. Experts and decision makers need to improve their awareness of how worries are the driving force of public engagement and that scientific knowledge rarely is an adequate response to these worries.

## Risk and Democracy: The Importance of Framing

How risks are defined is a central topic for sociology to study. The reason for this is that it determines what groups and what competences are considered relevant for taking a stand and making decisions (Lidskog et al. 2011). Sociology contributes to risk analysis by showing how definitions of risks shape social relations and distribute powers to groups, at the same time as other groups are excluded from decision making. In a risk context, a scientific definition of the issue at stake is often assumed. However, such a restriction may lead to reductionism, giving experts too much power, while rendering other important factors invisible. Laypeople are reduced to passive receivers of information who only can contribute their trust and consent regarding expert proposals (Wynne 1992b; Wynne 2001).

Sociological studies of expert work do not conclude that risk issues should be handled without experts. What is claimed is that experts alone should not define, investigate, and give answers to risk issues. Risks are too complicated to be delegated to experts. Sociological studies of the social dynamics in definition of risks are valuable for making risk management more relevant and robust. Not least to show how definitions and processes of framing are made and what consequences these processes have for different groups' possibilities to participate in and influence risk management.

Regulating a risk is not only about setting limits, but also about framing the risk as such, deciding what actors are of relevance and should be included in the decision process, what roles, mandates and responsibilities are given to them, and finally and most importantly what to make decisions about (Lidskog et al. 2009). Sociologists and other social scientists have critically scrutinized framing processes in public decision making, and a key finding is that frames concerning technical issues are usually dominated and influenced by experts in a technocratic way (Wynne 2001, 2005). By narrowing the issue, scientific experts exaggerate the scope, power, and importance of scientific knowledge in the public domain, neglecting cultural factors and ignoring citizen competence (Wynne 1992b; Wynne 2001; Jasianoff 2010). Paradoxically, science is often accorded the most prominent role even when public dialogue is striven for. A reason for this is that many issues are technically framed, with questions of risk, safety, and effectiveness placed at the center (Wynne 2005, 2010). Scientific expertise is needed

to answer such questions, since experts have the resources and competence to know the “true” nature of the issue at stake.

The result is that what is presented as a democratic risk decision process is often a technocratically framed process based on a scientific definition of the risk. The public are invited to participate in a process that is often presented as being about dialogue but the possibility to influence the process according to their own perspectives is often unclear, and in practice very restricted. These processes do not open up decision making to wider evaluation and influence, but instead function to gain legitimacy and acceptance for already defined – and many times in practice already decided – expert-based proposals.

A technocratic framing reduces the role of citizens to one of trusting or distrusting experts; to saying yes or no to already decided proposals and to being restricted to only discussing the local and concrete aspects of a project. Instead, they should be provided with opportunities to define and frame the project in their own way, putting forward what risks they see as relevant and worthy of attention. Sociologists have argued that the discussion should not only include the meaning of the project and its risk from the perspective of the experts and regulators, but also from the perspective of the public and other stakeholders (Gieryn 1999; Hilgartner 2000; Irwin and Michael 2003; Jasianoff 2005; Wynne 2005).

Since there is no correct framing of risk issues, risk management will always be surrounded by conflicts. Different groups frame problems differently and give them different priorities. Sometimes it is the case that what one group considers to be a solution to a problem, another group considers to be part of the problem. Some suggest that nuclear power is a sustainable energy solution, because it does not lead to carbon emissions, while others argue that the radioactive waste makes it anything but a sustainable and secure long-term source of energy. In this situation, the option of denying the existence of different frames is a dead end. Instead, the first step toward a robust solution is to acknowledge the existing frames and welcome different groups to contribute their own perspectives, using their own frames, knowledge, and values. Frequently, this opening up of the framing is met with criticism and opposition from those groups that are already handling the issue from a particular frame. They have invested time, money, and prestige in the project and are therefore reluctant to change the established framing of the issue and its particular way of handling it. But taking public involvement seriously entails a more democratic framing of issues, in the sense that issues have to be connected to public concerns (Marres 2007). If we are to find a way beyond pendulum swings between technocracy and populism – that either scientists or laypeople should decide – various groups of frames must meet on a more equal footing.

## Produced Risk: Beyond Realism and Constructivism

An ongoing controversy in risk research is between realism and social constructivism. Do risks possess physical characteristics that exist independently of cultural and social contexts, including actors' perceptions, or are they socially and culturally constructed attributes, produced and shaped by these contexts? Technical risk analysis is based on realism, and sees risks as independent of their context. As described earlier in this essay, many social scientists and sociologists accept technical risk analysis and its realism, but add studies on why the public accepts some risk analyses and rejects others. The public's risk assessments are understood as social constructs, whereas experts' risk assessments are seen as realistic descriptions. But among social scientists, we also find

those who question the realistic approach and consider it wrong. Instead, they want to go beyond this dichotomy in order to find a middle ground between realism and social constructivism. Many scholars claim that there is a third way between “naïve realism,” and “idealism” (Renn 2008), positivistic and constructivist paradigms (Rosa 1998), and “pure realism” and “radical constructivism” (Zinn 2008). The first implies the existence of an empirical and objective reality outside human perception, and the latter a subjective and cultural understanding shaped by humans and with no necessary connection to an objective reality.

However, the quest to find a middle way, or third way, between subjective and objective reality – between something internal and something external to human beings and society – reproduces what it tries to transcend. The first is based on causal laws of material reality, and the second, on a social world of opinions and norms. This point of departure is questioned by certain sociologists, claiming that the focus should instead be on the dynamic interplay between different factors that make up reality (Irwin and Michael 2003; Latour 1993, 2004, 2005). Reality is neither reducible to something out there, beyond human action, nor reducible to something in there, to human thoughts and actions. Instead it is co-produced by many factors.

However, social constructivism has been and still is an important tradition within the sociology of risk. Its historical roots go back to classical sociology, not least the work of Émile Durkheim (1858–1917). Durkheim elaborated a unique domain for sociology by demarcating a social reality totally different from that of biology and psychology. “The social” – or social facts, as he called it – is a reality in its own right, irreducible to other levels of reality (Durkheim 1982). The task of sociology was to explain social facts, and these explanations should not include any findings or factors from the psychological (individual) level or the biological level. The result was a specialization and division of labor among academic disciplines, where every discipline has its own domain and unique explanations. This understanding of sociology entailed that everything that exists outside the social domain was disregarded. The social domain was considered autonomous with regard to other domains.

The implication was that when analyzing risk, sociology should study people’s interpretations and experiences of these risks, and how these are bound to social structures that steer perceptions and actions. Material objects and technical artifacts were left outside the analysis, and seen as not having any power to influence what is taking place in the social domain. Experiences of risks should not be explained with reference to nature or artifacts, but only social factors. For example, when sociology explains people’s worries about nuclear waste, the focus should not be on the strength of the canisters as a technical barrier to protect the biosphere from radioactivity, but on people’s opinions about these barriers and how these influence their assessment of radioactive risks. Thus, the objects of sociological analysis are perceptions, interpretations, and socializations to social patterns of risk attitudes toward radioactivity and disposal of nuclear waste.

This sociological purification of a social dimension has been successful in so far as it has created a distinct niche for sociological thought and provided important knowledge concerning how people perceive, understand, and act upon risks. Nevertheless, its strong separation between nature and society, with sociology only investigating the latter, is problematic. This strong focus on the social dimension has led to a paradoxical understanding of nature and artifacts, which Bruno Latour (2004, p. 33) has aptly described as follows:

- ▶ Those who are proud of being social scientists because they are not naïve enough to believe in the existence of an “immediate access” to nature always recognize that there is the human history

of nature on the one hand, and on the other, the natural nonhistory of nature, made up of electrons, particles, raw, causal, objective things, completely indifferent to the first list.

The consequence of social constructivism is that we, on the one hand, find a society with a history and on the other a nature without history. Latour is critical of this kind of approach, which leaves important aspects of reality outside sociological analysis. According to him, the task of sociology is to transcend both realism and social constructivism, a task that necessarily entails that dichotomies – such as those between nature and culture, social and technical, actor and structure, science and society – be critically studied and not taken for granted. How and why these dichotomies are produced and reproduced should also be explained.

Latour's proposal for transcending the dichotomy between realism and social constructivism is to focus on the *production of risk*. Risks are produced by practices, by actors using instruments and technologies. It is therefore misleading as a sociologist to focus on perceptions, opinions, and experience. Instead, the focal point for sociology should be to explore how risks are produced, by what means, and with what effects. The focus on practices means that there is no “real risk” behind our perceptions and actions. There are no risks separate from actors and society, possible to observe by actors. Instead, there are a number of actors and activities where nature, technology, and culture interact, resulting in the production of risks. There are no risks beyond socially produced risks, that is, beyond the measuring and monitoring of risk. Through these practices not only knowledge about risks is produced but also the risks as such. Thus, practices are performative; they not only describe reality but also shape it. By studying these practices, sociology can transcend the dichotomy between realism and social constructivism.

## **Governmentality: Toward an Individualized Risk Management**

Ulrich Beck emphasizes that the current society is increasingly individualized, in the sense that individuals are seen as being responsible creators of their own lives and are therefore constantly required to make their own decisions. “The choosing, deciding, shaping human being who aspires to be the author of his or her own life, the creator of an individual identity, is the central character of our time,” as Beck (2002, p. 23) puts it.

This individualization, however, does not necessarily mean the achievement of greater personal freedom. Beck grasps this development with the term “institutionalized individualism” (Beck and Beck-Gernsheim 2002). At the same time, as nation-states have outsourced many of their functions and operations, there is an insourcing of functions to the individual level. What the nation-state, the employer, the union, or the family once provided is now presented as being the responsibility of the individual. Thus, individualization in this sense does not mean freedom of choice, but instead the compulsion to choose in a situation where no certainties exist. It is a “precarious freedom” centered on imperatives such as think, calculate, plan, adjust, negotiate, define, and revoke (Beck 2002). But even though we often lack knowledge of what choices are best, it is demanded of us to make individual decisions and be responsible for the consequences.

There is a tension between institutionalized individualism and the risk society thesis about mega-hazards beyond human control. According to Beck (2008), the government of incalculable risks and mega-hazards leads to the irony of putting an end to the free liberal society in the ambition of protecting citizens from risks. At the same time, individuals are continuously ascribed responsibility for risks that are impossible for them to manage.

However, a certain strand of sociological thought, following Michel Foucault's (1991) work on governmentality, cultivates a perspective that takes neither individualism nor the character of risks and the risk society for granted. Instead they argue that risks should be conceptualized and understood as a way of steering practice. The task of sociology is to study how, through technical apparatus and administrative institutions, incalculable dangers are made into knowable and governable risks. Risks become a way of ordering reality and making it calculable, and expert knowledge is decisive in this (Rose 1993; Dean 1999).

Instead of making use of coercive power, the government can steer through norms, knowledge, and individual self-discipline. The reason for this is that we have today an "advanced liberal society," based on a clear division between the state and the civil society (Rose 1996, 1999). Civil society has emerged as an autonomous sphere in which individuals can express themselves as free citizens. In protecting this autonomy, coercive means of governmental control are precluded, which means that more sophisticated instruments and mechanisms need to be developed, technologies for governing at a distance (Rose and Miller 1992).

This way of exercising power, in which those who are controlled feel autonomous, is based on tools for the self-development of those who are governed. The responsibility is placed on citizens to govern themselves, to act upon themselves, and be responsible "for the security of their property and their persons, and that of their families" (Rose 1999, p. 247). An almost paradoxical relationship is created between the state and the civil society, in which the exercise of power is conducted with the goal of not being visible. It is characterized more by bringing citizens to perform a regulated freedom than by imposing on them coercive measures (Rose and Miller 1992, p. 174).

The strong emphasis on individuals as being responsible governors of their own lives creates dilemmas. Increasingly, individuals have to face and make decisions on a range of issues characterized by uncertainty. An example of this is how genetic risks are governed with the aim to improve the quality of the population. During the development of the welfare state, it became an important task for the government and the public administration to guide, control, and intervene in the reproduction of the population, but today these decisions are delegated to individual citizens. The problem is no longer framed as improving the quality of the population but as a question of individual lifestyles. Today, reproduction is about promoting the self-governance of the client (Novas and Rose 2000). The responsibility to govern genetic risks – to decide about having children and informing others about one's own genetic risks – has been made into a lifestyle choice. However, plenty of experts are willing to give advice on how to make your own lifestyle possible, and guide your choice in certain directions.

Risks thereby constitute a strategy for disciplinary power to monitor and govern individuals and thereby whole populations (O'Malley 2008). Those individuals that deviate from what is presented as normal behavior are seen as "at risk," and need to be controlled with the aim of achieving behavioral modification. This control is primarily that of self-management, with individuals being urged to protect themselves from certain risks (Giddens 1991). Risks are thereby de-socialized, privatized, and individualized; they become a responsibility of the individual, and a way for government to govern the conduct of individuals. The sociology of risk should therefore be devoted to studying questions about how problems are defined, by whom and in relation to what goals, and through which practices, technologies and rationalities this governing is accomplished and authority exercised.

## Further Research

---

As emphasized in the introduction, the specific contribution of sociology of risk is to place risk in its social context. There are no risks “out there” in the sense of being independent of the society in which they emerge, are measured and monitored. Society is differentiated, which means that cognitions, understandings, and feelings of risks are differentiated. Actors – including scientific ones – have various structural positions and cultural belongings and therefore understand risks differently. To develop sociological knowledge on risks implies to contextualize risks; to associate them with specific actors, institutions, and settings. This means that no conceptualization, regulation, or research on risks is beyond sociological exploration; and furthermore, scientific definitions of risks and technical risk analysis should be proper study objects for sociological investigation.

This does not imply a reductionism and relativism, seeing different actors’ understandings of risks as all that exists. On the contrary, risks should be understood as produced through social activities where nature, technology, and culture interact. Sociology of risk should not be restricted to investigating risk perceptions, but should also study definitions and usage of risks, including how different actors deal with risky nature and unruly technologies; how these are framed and regulated, and as a consequence of these activities, produced.

The five thematic areas described above have by no means been given a final answer, but are in need of further research. As already emphasized, these areas are interrelated; organizational aspects of governing risks, public inclusion in risk regulation, framing and production of risks, and the monitoring of individuals’ risk behavior are interconnected. As with many other disciplines, sociology consists of different theoretical traditions, methodological assumptions, and analytical approaches. Therefore, it will never be able to give simple, single, and final answers to complex issues, and the sociology of risk is no exception of this. It is, however, able to gain knowledge on important topics, and while not producing final knowledge at least it may produce more and better knowledge – theoretically informed and empirically sensitive – on the function and place of risks in different social settings.

Studying processes of risk assessment and risk management, the sociology of risk could make important contributions to preparing and realizing a political and democratic discussion on risk issues, controversial as well as uncontroversial. By identifying and clarifying the political aspects of these objects, making frames and framing processes visible, and showing how technologies and political devices are embedded in social processes, it opens up risk regulatory processes for public scrutiny and evaluation. What may originally be framed as technical issues, only relevant for a specialized group of experts, will thereby become relevant for citizens. Risk regulation is about more than just choosing the best regulatory instruments and finding the best technical solutions to predefined risks. It concerns building society and choosing a future.

---

## References

---

- Amendola A (2001) Recent paradigms for risk informed decision making. *Safety Sci* 40(1):17–30
- Baldwin R, Cave M (1999) Understanding regulation. Oxford University Press, Oxford
- Bauman Z (1993) Postmodern ethics. Blackwell, Oxford
- Bauman Z (2006) Liquid fear. Polity, Cambridge
- Beck U (1992) Risk society. Towards a new modernity. Sage, London

- Beck U (2002) A life of one's own in a runaway world: individualization, globalization and politics. In: Beck U, Beck-Gernsheim E (eds) Individualization, institutionalized individualism and its social and political consequences. Sage, London, pp 22–29
- Beck U (2008) Living in the world risk society. *Econ Soc* 35:329–345
- Beck U, Beck-Gernsheim E (2002) Losing the traditional. Individualization and “precarious freedom”. In: Beck U, Beck-Gernsheim E (eds) Individualization. Institutionalized individualism and its social and political consequences. Sage, London, pp 1–21
- Bickerstaff K, Walker G (2001) Public understandings of air pollution: the “localization” of environmental risk. *Glob Environ Chang* 11:133–145
- Bloor D (1976) Knowledge and social imagery. Routledge, London
- Breakwell GM (2007) The psychology of risk. Cambridge University Press, Cambridge
- Calhoun C, Gerteis J, Moody J, Pfaff S, Virk I (2007) Contemporary sociological theory, 2nd edn. Blackwell, Oxford
- Callon M, Lascombes P, Barthe Y (2009) Acting in an uncertain world: an essay on technical democracy. MIT Press, Cambridge, MA
- Dean M (1999) Risk, calculable and incalculable. In: Lupton D (ed) Risk and sociocultural theory. Cambridge University Press, Cambridge, pp 131–159
- Douglas M (1966) Purity and danger. An analysis of the concepts of pollution and taboo. Routledge Kegan Paul, London
- Douglas M (1978) Cultural bias. Royal anthropological institute of Great Britain and Ireland, London
- Douglas M (1982) Introduction to grid/group analysis. In: Douglas M (ed) Essays in the sociology of perception. Routledge Kegan Paul, London, pp 1–8
- Douglas M (1992) Risk and blame: essays in cultural theory. Routledge, London
- Douglas M (1996) Thought styles: critical essays on good taste. Sage, London
- Douglas M, Wildavsky A (1982) Risk and culture: an essay on the selection of technological and environmental dangers. University of California Press, Berkeley
- Durkheim E (1982) The rules of sociological method. Macmillan Press, New York
- Eriksson B (1993) The first formulation of sociology. A discursive innovation of the 18th century. *Eur J Sociol* 34(2):251–276
- Ferguson A (1767/1996) Essay on the history of civil society. Cambridge University Press, New York
- Finucane M, Slovic P, Mertz CK, Flynn J, Satterfield T (2000) Gender, race and perceived risk: the “white male” effect. *Health Risk Soc* 2(2):159–172
- Foucault M (1991) Governmentality. In: Burchell G, Gordon C, Miller P (eds) The Foucault effect: studies in governmentality. Harvester Wheatsheaf, Hemel Hempstead, pp 87–104
- Funtowicz SO, Ravetz JR (1990) Uncertainty and quality in science for policy. Kluwer, Dordrecht
- Furedi F (2002) The culture of fear. Risk-taking and the morality of low expectation. Continuum, London
- Furedi F (2008) Fear and security: a vulnerability-led policy response. *Soc policy adm* 42(6):645–661
- Giddens A (1984) The constitution of society. Outline of the theory of structuration. Polity, Cambridge
- Giddens A (1990) The consequences of modernity. Polity, Cambridge
- Giddens A (1991) Modernity and self-identity. Polity, Cambridge
- Giddens A (1994) Risk, trust and reflexivity. In: Beck U, Giddens A, Lash S (eds) Reflexive modernization. Politics, tradition and aesthetics in the modern social order. Polity, Cambridge, pp 184–197
- Giddens A (1999) Runaway world. How globalization is reshaping our lives. Routledge, New York
- Gieryn TF (1999) Cultural boundaries of science: credibility on the line. University of Chicago Press, Chicago
- Gouldson A, Lidskog R, Wester-Herber M (2007) The battle for hearts and minds. Evolutions in organisational approaches to environmental risk communication. *Environ Plann C* 25(1):56–72
- Gutteling J, Wiegman O (1996) Exploring risk communication. Kluwer, Dordrecht
- Hilgartner S (1992) The social construction of risk objects: or, how to pry open networks of risks. In: Short JF, Clarke L (eds) Organizations, uncertainties, and risk. Westview Press, Boulder, pp 39–53
- Hilgartner S (2000) Science on stage: expert advice as public drama. Stanford University Press, Stanford
- Hobbes T (1651/2005) Leviathan. Continuum, London
- Höijer B, Lidskog R, Uggla Y (2006) Facing dilemmas. Sense-making and decision-making in late modernity. *Futures* 38(3):350–366
- Hood C, Rothstein H, Baldwin R (2001) The government of risk. Understanding risk regulation regimes. Oxford University Press, Oxford
- Howel D, Moffatt S, Prince H, Bush J, Dunn C (2002) Urban air quality in North-East England: exploring the influences on local views and perceptions. *Risk Anal* 22(1):121–130
- Hughes E, Kitzinger J, Murdock G (2006) The media and risk. In: Taylor-Gooby P, Zinn J (eds) Risk in social sciences. Oxford University Press, Oxford, pp 250–270
- Hutter BM (2001) Regulation and risk. Occupational health and safety on the railways. Oxford University Press, Oxford

- Hutter B, Power M (eds) (2005) *Organizational encounters with risk*. Cambridge University Press, Cambridge
- Irwin A (2006) The politics of talk. Coming to terms with the “new” scientific governance. *Soc Stud Sci* 36(2):299–320
- Irwin A, Michael M (2003) *Science, social theory and public knowledge*. Open University Press, Maidenhead
- Irwin A, Wynne B (eds) (1996) *Misunderstanding science? The public reconstruction of science and technology*. Cambridge University Press, Cambridge
- Jaeger CC, Renn O, Rosa EA, Webler T (2001) *Risk, uncertainty and rational action*. Earthscan, London
- Japp KP, Kusche I (2008) System theory and risk. In: Zinn J (ed) *Social theories of risk and uncertainty. An introduction*. Blackwell, Oxford, pp 76–105
- Jasanoff S (ed) (2004) *States of knowledge: the co-production of science and social order*. Routledge, London
- Jasanoff S (2005) *Designs of nature. Science and democracy in Europe and the United States*. Princeton University Press, Princeton
- Jasanoff S (2010) A new climate for society. *Theory culture soc* 27(2–3):233–253
- Kasperson RE (1992) The social amplification of risk: progress in developing an integrative framework of risk. In: Krinsky S, Golding D (eds) *Social theories of risk*. Praeger, Westport
- Kasperson RE, Renn O, Slovic P, Brown HS, Emel J, Goble R, Kasperson JX, Ratick S (1988) The social amplification of risk: a conceptual framework. *Risk Anal* 8(2):177–187
- Kasperson JX, Kasperson RE, Pidgeon N, Slovic P (2003) The social amplification of risk: assessing fifteen years of research and theory. In: Pidgeon N, Kasperson RE, Slovic P (eds) *Social amplification of risk*. Cambridge University Press, Cambridge, pp 13–46
- Kemshall H (2006) Crime and risk. In: Taylor-Gooby P, Zinn J (eds) *Risk in social sciences*. Oxford University Press, Oxford, pp 76–93
- Krinsky S, Golding D (eds) (1992) *Social theories of risk*. Praeger, Westport
- Latour B (1993) *We have never been modern*. Harvester Wheatsheaf, New York
- Latour B (2004) *Politics of nature: how to bring the sciences into democracy*. Harvard University Press, Cambridge, MA
- Latour B (2005) *Reassembling the social: an introduction to actor-network theory*. Oxford University Press, Oxford
- Levinson R, Thomas J (eds) (1997) *Science today: problem or crisis?* Routledge, London
- Lidskog R (1996) In science we trust? on the relation between scientific knowledge, risk consciousness and public trust. *Acta Sociologica* 39(1):31–56
- Lidskog R (2008) Scientised citizens and democratised science. Re-assessing the expert-lay divide. *J risk res* 11(1–2):69–86
- Lidskog R (2011) Regulating nature: public understanding and moral reasoning. *Nature and Culture* 6(2):149–167
- Lidskog R, Sundqvist G (2011) The science-policy-citizen dynamics in international environmental governance. In: Lidskog R, Sundqvist G (eds) *Governing the air: science-policy-citizen dynamics in international environmental governance*. MIT Press, Cambridge, MA, pp 323–359
- Lidskog R, Soneryd L, Uggla Y (2005) Knowledge, power and control: studying environmental regulation in late modernity. *J Environ Policy Plann* 7(2):89–106
- Lidskog R, Soneryd L, Uggla Y (2009) *Transboundary risk governance*. Earthscan, London
- Lidskog R, Uggla Y, Soneryd L (2011) Making transboundary risks governable: reducing complexity, constructing identities and ascribing capabilities. *Ambio* 40(2):111–120
- Löfstedt RE (2005) *Risk management in post-trust societies*. Palgrave Macmillan, New York
- Luhmann N (1984) *Soziale systeme. Grundriss einer allgemeinen theorie*. Suhrkamp, Frankfurt a M
- Luhmann N (1989) *Ecological communication*. Polity, Cambridge
- Luhmann N (1993) *Risk. A sociological theory*. Walter de Gruyter, Berlin
- Lupton D (1999) Introduction: risk and sociocultural theory. In: Lupton D (ed) *Risk and sociocultural theory*. Cambridge University Press, Cambridge, pp 1–11
- Marres N (2007) The issues deserve more credit: pragmatist contributions to the study of public involvement in controversy. *Soc Stud Sci* 37(5):759–778
- Millar J (1771) *Observations concerning the distinctions of rank in society*. T Ewing, Edinburgh
- Novas C, Rose N (2000) Genetic risk and the birth of the somatic individual. *Econ Soc* 29(4):485–513
- O’Malley P (2008) Risk and governmentality. In: Zinn JO (ed) *Social theories of risk and uncertainty. An introduction*. Blackwell, Oxford, pp 52–75
- Pidgeon N, Kasperson RE, Slovic P (eds) (2003) *Social amplification of risk*. Cambridge University Press, Cambridge
- Power M (2007) *Organized uncertainty. Designing a world of risk management*. Oxford University Press, Oxford
- Reith G (2009) Uncertain times: the notion of “risk” and the development of modernity. In: Löfstedt RE,

- Boholm Å (eds) *The Earthscan reader on risk*. Earthscan, London, pp 53–68
- Renn O (1998) Three decades of risk research: accomplishments and new challenges. *J Risk Res* 1(1):49–71
- Renn O (2008) Risk governance. Coping with uncertainty in a complex world. Earthscan, London
- Rosa EA (1998) Metatheoretical foundations for post-normal risk. *J Risk Res* 1(1):15–44
- Rose N (1993) Government, authority and expertise in advanced liberalism. *Econ Soc* 22:283–299
- Rose N (1996) Governing “advanced” liberal democracies. In: Barry A, Osborne T, Rose N (eds) *Foucault and political reason*. UCL Press, London, pp 37–64
- Rose N (1999) Powers of freedom: reframing political thought. Cambridge University Press, Cambridge
- Rose N, Miller P (1992) Political power beyond the states: problematics of government. *Br J Sociol* 43(2): 173–205
- Sjöberg L (2000) Perceived risk and tampering with nature. *J Risk Res* 3(4):353–367
- Slovic P (1987) Perceptions of risk. *Science* 236:280–285
- Slovic P, Peters E (1998) The importance of worldviews in risk perception. *Risk Decis Policy* 3(2):165–170
- Smith A (1776/2001) *An inquiry into the nature and causes of the wealth of nation*. Random House International, New York
- Starr C (1969) Social benefit versus technological risk. *Science* 165(3899):1232–1238
- Stehr N (2005) Knowledge politics. Governing the consequences of science and technology. Paradigm, Boulder
- Sundqvist G, Elam M (2010) Public involvement designed to circumvent public concern? The ‘participatory turn’ in European nuclear activities. *Risk Hazards Crisis Public Policy* 1(4):203–229, Article 8
- Taylor-Gooby P, Zinn J (eds) (2006) *Risk in social sciences*. Oxford University Press, Oxford
- Thompson M, Ellis R, Wildavsky A (1990) Cultural theory. Westview Press, Boulder
- Todt O (2003) Designing trust. *Futures* 35:239–251
- Tulloch J (1999) Fear of crime and the media: sociocultural theories of risk. In: Lupton D (ed) *Risk and sociocultural theory*. Cambridge University Press, Cambridge, pp 34–58
- Tulloch J, Lupton D (2003) *Risk and everyday life*. Sage, London
- Wester-Herber M (2004) Underlying concerns in land-use conflict: the role of place-identity in risk perception. *Environ Sci Policy* 7(2):109–116
- Wynne B (1992a) Misunderstood misunderstanding: social identities and public uptake of science. *Public Underst Sci* 1(3):281–304
- Wynne B (1992b) Uncertainty and environmental learning: reconceiving science and policy in the preventive paradigm. *Glob Environ Chang* 2(2):111–127
- Wynne B (1993) Public uptake of science: a case for institutional reflexivity. *Public Underst Sci* 2(4):321–337
- Wynne B (1995) Public understanding of science. In: Jasianoff S, Markle GE, Peterson JC, Pinch T (eds) *Handbook of science and technology studies*. Sage, Thousand Oaks, pp 361–388
- Wynne B (1996) May the sheep safely graze? A reflexive view of the expert-lay knowledge divide. In: Lash S, Szerszynsky B, Wynne B (eds) *Risk, environment modernity. Towards a new ecology*. Sage, London, pp 44–83
- Wynne B (2001) Creating public alienation. Expert cultures of risk and ethics on GMOs. *Sci cult* 10(4):445–481
- Wynne B (2005) Risk as globalizing ‘democratic’ discourse? Framing subjects and citizens. In: Leach M, Scoones I, Wynne B (eds) *Science and citizens. Globalization and the challenge of engagement*. ZED Books, London, pp 66–82
- Wynne B (2010) Strange weather, again. Climate science as political art. *Theor Cult Soc* 27(2–3):289–305
- Zinn JO (2008) A comparison of sociological theorizing on risk and uncertainty. In: Zinn JO (ed) *Social theories of risk and uncertainty. An introduction*. Blackwell, Oxford, pp 168–210
- Zinn JO, Gooby-Taylor P (2006) Risk as an interdisciplinary research area. In: Taylor-Gooby P, Zinn J (eds) *Risk in social science*. Oxford University Press, Oxford



# 41 Risk and Gender: Daredevils and Eco-Angels

Misse Wester

The Royal Institute of Technology, Stockholm, Sweden

<i>Introduction</i> .....	<b>1030</b>
<i>History</i> .....	<b>1031</b>
<i>Current Research</i> .....	<b>1035</b>
Knowledge of and Familiarity with Science .....	1036
Issues of Trust .....	1037
Differences in Risk Perception due to Biological or Social Differences .....	1038
Cultural Explanations .....	1042
Cultural Divisions: Practical Implications .....	1043
Summary .....	1044
<i>Further Research</i> .....	<b>1045</b>

**Abstract:** Through history and in most cultures differences between men and women have been observed and expressed in various ways. Also in our society differences between men and women are evident and this becomes particularly clear in the context of risk perception. This chapter will present three models that are found in the research literature that aim at explaining these differences. The first model focuses on differences due to knowledge and familiarity with science, including trust. This model however, fails to take into consideration the distinction between estimated knowledge and factual knowledge, making it difficult to understand exactly what role knowledge plays in risk perception. Second, differences between men and women have also been explained by biological mechanisms or social roles. Here it is believed that women are more nurturing by nature and men are more driven toward growth and expansion. It is argued that it is of limited importance whether these differences are biological or social, as it will have little bearing on the management of risk. The third explanation focuses on cultural differences and uses examples from disaster management. Here it is concluded that women are by social processes most often excluded from the areas where risks are created and managed, and suffer the consequences from this lack of influence in disasters. The chapter concludes with three suggestions for future research.

## Introduction

---

How risks are perceived depend on a number of things: for example, how old we are, what experiences and education we have, and if the risk is taken voluntary or not. One way of looking at the perception of risk is to focus on specific characteristics of the risk itself: if it is a novel risk or a risk that is perceived as unfair. Also the consequences of the risk influence risk perception: if the consequences are irreversible rather than temporary, chances are measures to strictly regulate this risk will be taken. Our individual differences and preferences also affect how we perceive risks: which risks are worth taking, which are not, and how risks should be managed are all affected by individual differences. These variations in risk perception can be due to differences in age or education, personal experiences or nationality. In many contexts, differences between men and women how risks are perceived and acted upon are found. For example, it is a common result that men score lower on risk perception scales than women. Also, women express more concern for the environment and state to be more willing to take steps in order to improve the environment than men do. Some theories have been suggested to explain these differences. Some state that women are by nature more nurturing than men, making them biologically wired to be more risk averse. However, the differences between men and women are not consistent in all studies of risk. This lack of coherence suggests that either differences between men and women are not stable or predictable, or that questions relating to risk are so complex that simple dichotomies between people are not good explanations. Other theories suggest that differences between men and women are cultural, rather than biological, and that gender differences should be examined from a perspective that includes issues of power, stereotypes, and influence. In this chapter, I will focus on differences between men and women found in studies concerning the perception of technical and environmental risks. These differences are most often observed although not always explained, but a few models that have attempted to explain these differences will be discussed. The chapter continues with a discussion of current research and ends with some suggestions on future directions.

Before moving on, I will attempt to position myself on the risk field. In most contexts, there is a division between different types of risks. There are those risks that we are exposed to on an everyday basis. These types of risks include getting hit by a car while crossing the street or falling off a chair while changing a lightbulb. Close to this category are the risks that we face as a result of individual choices in lifestyle for example smoking or making specific dietary choices, or engaging in more risk-filled activities such as driving recklessly, skydiving, and base jumping. Differences between men and women in this category of risks – both with regard to perception and behavior – will not be addressed at any length in this chapter, even though gender differences are observed here (see, e.g., DeJoy 1992; Andersson and Lundborg 2007; Deeks et al. 2009). Instead, this chapter will focus primarily on risks that are imposed on us, where the options for personal control and opting-out are limited. Yet another line that divides different risks is the cause or origin of the risk itself and this can also affect perception. For example, even if the consequences are the same, such as number of injured or exposure to chemicals, the reactions will be vastly different if the risk is caused by an unforeseen accident or because of neglect (see Weisæth et al. 2002; Wester 2009). In this chapter, this will not be of specific focus, even if the effect of cause or origin of risk and how this links to gender still needs further research (Finucane et al. 2000). The concept of risk will be used in a very wide context with the full understanding that risk is a multifaceted and complex concept. However, the main focus here is to account for differences in perception between men and women, and what implications this has for risk management.

## History

---

The distinction between men and women as fundamentally different beings is certainly not a novel concept. Early records reveal that Aristotle suggested that a woman actually is a man that had failed to develop to his full potential (Horowitz 1976; Merchant 1980). This division between men and women was not just a mere observation but also included values: masculine characteristics were more highly valued than feminine attributes and traits. In Aristotelian time, free men were regarded as superior to women and slaves, thereby consolidating the differences and values associated with them (Horowitz 1976). Treating men and women differently is not unique to western societies. Anthropological studies have revealed that men and women have not only been perceived as distinct from each other in other cultures, but have also lived separately. With the first ethnographic field studies in the nineteenth century, ethnographers discovered that in some parts of Papua New Guinea, for example, young boys are moved from their mothers home around the age of seven and moved into the men's house. In this house, boys would grow into full men under the supervision of other, older, and initiated men and into this house women were not allowed (Keesing 1981; Kulick 1987). This division between men and women is not limited to them living in different areas; it has also lead to a division of labor in most cultures. Caring for children and performing domestic chores and small-scale agriculture have traditionally been the domain of women, whereas men have primarily performed their duties outside of the home. In our Western society, differences in distribution of men and women in different areas are clearly visible. For example, science and technology have long been areas where men have outnumbered women, whereas occupations in the service or public sector, female employees are in majority. This separation between men and women in the workforce corresponds with unequal representation in various

fields of education. This division of men and women is found in many levels of society such as differences in academic degrees, differences in wages, and an underrepresentation of women in executive positions. Given this division of gender, it is not surprising that there will be differences between how men and women perceive and relate to the world.

Still today, the view of men and women as different from each other is apparent in many contexts. Even if it is recognized that women and men are not entirely different species, popular books like “Men are from Mars, Women are from Venus” make it clear that there are differences between sexes that need explaining. The word processing program used for composing this contribution defines “female” as: feminine, motherly, and soft, whereas “masculine” is defined as: virile, manly, brave, and proud. Even though there has been a shift in language from “sex differences” to “gender differences,” implying that focus is not so much on biology as it is on social role, men and women are still perceived as separate from each other. No longer reduced to biological creatures, the function of estrogen and testosterone is still widely held to affect not only the physical development of sexes, but also to influence preferences or specific behavior. However, it is becoming more widely acknowledged that specific behavior relates as much to stereotypical roles as they do to biology. Early in childhood, we learn what it means to be a boy and to be a girl. How these stereotypes are internalized and develop are of course complex matters, but research suggests that children as young as 3 years old realize that men and women are different, and some suggest that as early as between 2 and 4 years of age, it is clear to both girls and boys what is expected of them and others as gender-specific behavior (Bem 1983; Weintraub et al. 1984). Even though the concepts of what is male and female differs over time and within cultures, it might be safe to say that most of us do incorporate images of what it entails to be a man or a woman into our way of thinking about the world and how it works. This also means that we learn how to behave in accordance with the gender roles – whether we agree with them or not. Within social psychology, most scholars agree that differences between men and women in many areas are the result of placing men and women in stereotypic roles and ascribing typical attributes to individuals based on their gender (Eisler et al. 2003). Now, in order to navigate a complex world as the one we live in, there are situations where relying on stereotypes makes life a bit easier. For example, when we meet a new person we rely on known rules of engagement for that situation: we shake hands and introduce ourselves in a given order. There are also situations where we encounter groups of people; like students, tourists, or costumers, we expect these encounters to follow a certain pattern. Any breaks with the norm can cause us to reevaluate something taken for granted, which leads to us having to change our preconceived ideas and question our perception of the world. The implication of this is that we often have a tendency to rely on stereotypes and to do what is expected of us. An obvious question is how this internalization of specific roles affects us throughout our lives, what education we choose, what we expect from others, and how these differences are linked to how we perceive the world. The focus in this chapter is how the differences in gender affect risk issues. More specifically, this chapter will focus on what implications differences between men and women have for shaping risk perception and what measures are taken in order to address some of the risks that we face. For this purpose, the role of potential variations due to biological differences is not important. For the purpose of the argument presented here, whether the differences between men and women are caused by biology or social processes is subordinate to the discussion of what the implications of these differences are. In other words, it does not matter if testosterone makes men behave in a certain way or if this behavior that we associate with the male sex is learned behavior. This is because in

the context of risk perception and risk management, using biological differences as a reason for prioritizing the view of one group over another is not a legitimate reason. In this chapter, issues that influence the perception and acceptance of genetically modified organisms (GMOs) or measures taken in order to adapt to climate change and the levels of testosterone or estrogen in individuals are of limited interest.

In an overwhelming majority of empirical studies on matters ranging from environmental risk perception to risk-taking activities, there are systematic and consistent differences between men and women. In this context, one legitimate question one might ask is “so what?” What does it matter if we think that women are more emotional than men; that boys are louder and fight more than girls; or that women are biologically wired to care for children? If we are free to decide on the content and direction of our own lives, the stereotypes will not affect me or my family – so what? This attitude could be true if we lived our lives in isolation or a social vacuum, but most of us do not. When differences between men and women are expressed systematically, it means that men and women have different views of risks in a general sense and this is not a result of a single study or single fluctuation. Systematic differences between any two or more groups are results of a stable and reoccurring factor that affects both groups to an extent that is measurable over time. This includes differences in opinion on which risks are perceived to be acceptable, how benefits are valued, what risks are to be avoided, and how resources should be allocated in order to mitigate specific risks. In these cases, we have to choose one view over another and act in line with that decision. As the review of research literature and discussion in this chapter hopes to illustrate, is that the dominant view of risks is more aligned with the perception that men hold rather than the view that women have, and by choosing this we run into issues that need to be addressed. Of course, choices need to be made and priorities set, but in a democratic society, one groups’ perception cannot systematically be the dominating one. Just like the biological claims of racism are acknowledged to be unfounded, and discrimination against individuals because of race or ethnic belonging is not accepted, so is sexism recognized to be unjust and morally questionable. This is not the same as saying that there are no differences between men and women, or that men and women as groups are internally homogenous. However, there might just as well be differences between tall and short individuals, or between individuals that prefer yellow cars to green cars, but these are questions we rarely ask.

Sociologist Ulrich Beck claims that we live in a risk society (Beck 1986). In the risk society, citizens are faced with risks that are produced as a direct of modern technology and these risks often elude our senses. As most individuals do not have the means to detect, for example, radiation or the presence of hazardous chemicals in our food or water by using our sight, smell, taste, or touch, these “new” and invisible risks that ordinary citizens are exposed to but cannot detect are heavily dependent on science, both for the detection of and management of risks. This means that risks become known after an adverse effect has already been identified, that individuals have already become ill from eating contaminated food, or that we can see the negative effects in nature. If we accept that we live in a risk society, differences between men and women cannot be fully understood unless risks are seen in this broader context. Some would argue that risks are “out there” in the natural environment for the scientific community to discover. Other scholars argue that risks are socially constructed and contain as much social elements as technical ones. Whatever perspective one chooses, and perhaps a middle ground is to be preferred, the first step in discovering (or creating) a risk as the risk identification. Once we have identified a risk, a hazard assessment needs to be done. In this step focus is on

what the danger is, who or what is exposed to the risk, and what consequences this can lead to. After this, policy decisions have to be made and priorities have to be set. Risks are not distributed fairly and as a society we make decisions on what risks we are willing to accept and what risks are to be avoided at all costs. The process of identifying and deciding what risks to avoid has traditionally been a technical task that requires knowledge in technology and the natural sciences. Compared to 150 years ago, the risks we face today are vastly different. There are risks that we cannot see, taste, or smell, and in order to detect these risks, special tools and knowledge are needed. The consequences are also different today: One failed crop for a family in a small community would strike hard at that family; one failed crop for a multinational actor can have effect on the international price of grain on the global market. If we accept that risk society has made us dependent on experts that possess the knowledge and have access to the tools needed to identify and assess the risks we face, then we must take a closer look at who are included in this group and examine their risk perception. The process of risk management can in short be seen like this: there are risks in our society that are known but also some that are unknown. In order to determine what risks we need the most protection from, some priorities have to be made. This setting of priorities is the first instance where we have to choose what to protect. One example of this can be a choice between prioritizing human health or the environment, where different values or preferences act to guide our decision. As can be seen in controversies over certain risk issues, environmentalist groups can have a different set of values than the local environmental protection agency and this might cause conflict between the two views. Once a risk has been identified as something that needs further investigation, the process of risk assessment starts. Here, what is dangerous is determined and at what levels. In this process, there is a need for experts in various fields that will be able to contribute specific knowledge on a variety of issues including hazard identification, dose-response ratios, and calculating exposure scenarios. These steps of the process are fairly technical and the individuals that have access to the information in this stage of the process are often scientists working in labs. Once the risk has been assessed, now comes the time to decide how this risk is best managed. In this stage, the risk has left the lab and is increasingly becoming an issue for regulators and policy makers. In this stage, questions on how best to avoid, reduce, or accept some of the negative possibilities with the risk arise. When the risk management has reached this stage, there can also be great differences in how different groups view what resources should be put toward certain risk mitigation and what constituted an acceptable risk. In this context, it becomes clear that the risk identification process and the decision on the distribution of resources are related to issues of power and political influence (see Slovic 1999). Looking at what individuals that are involved in various stages of the process, there is a need for natural scientists, engineers, and medical experts. At most universities, women are a minority as students of engineering and natural sciences and this creates an unbalance in the distribution of gender in the expert groups. In the risk society, the experts that are called upon to make risk identifications, risk assessment, and handle risk management are predominately male. But the differences do not stop at this end of the risk management spectrum. In most risk research studies, women express a higher risk perception than men and also report being more willing to take pro-environmental action, indicating that men and women not only have separate views on risks, but also choose different ways of dealing with risk issues. Differences in perceptions of risk between men and women are observed in the majority of studies, but the reasons for these differences are rarely discussed and the implications of this discrepancy on a management level have received limited attention.

In this chapter, these issues are among those that will be expanded upon. The source or cause of the risk treated here are those risks that are dependent on or created by science and technology, where environmental risks are included.

## Current Research

---

Risks are central to our lives. It can be argued that without risks, there can be no growth. When we subject ourselves to risks, it is done in order to gain or benefit something. As societies have changed, so have the stakes. The risks we create today can have consequences that are not immediately visible, that might manifest in an unforeseen place and manifest in way so the consequences are irreversible.

In the field of risk perception, social scientists have worked for many years to increase our understanding on how different groups in society view risks. Some would say that as our society has grew more and more complex and the identification and management of risks issues have become primarily a scientific task, the differences in opinion between different groups have become more apparent. It might be important to point out that risk perception does not translate into fear or worry. Instead, a high-risk perception can be interpreted as a strong reaction toward a risk, a stronger desire to have that risk regulated or an increased concern over possible consequences of a risk. Risk perception does not necessarily measure an emotional reaction to a risk, even if there are examples of this. Instead, in this context I would like the reader to interpret the term “risk perception” as an opinion or reaction toward a risk. This reaction can be based on many things; such as moral considerations, personal control, ability to take action, or demands for stronger regulation, or outrage, but fear or worry is not the primary experience. Today it is more the rule than an exception that risks are debated and differences in opinion are voiced. There are many factors that influence the way we perceive risks. Some factors that influence risk perception of technical, or man-made risks, can be said to be perceived benefits, acceptance of a technology, and trust in the institutions that regulate the risk (Davidson and Freudenburg 1996; Siegrist 2000). Age is one example, where differences between generations are found on most risk issues. In some cases the younger generation is more risk averse, whereas the older generation sees a risk as less threatening, and in other situations it is the other way around. Education and experiences also influence how we view risks, and linked to this are values and ideologies. For example, if a student chooses to study nuclear physics, chances are that this student will pursue a professional career in the nuclear energy field. It is unlikely, but not unheard of, that a person will spend years in training to a substantial financial cost, only to end up working in a field that deals with a form of energy production that this person opposes. In this regard, education is linked to values, and in some cases values are reflected in ideology. Continuing with the example of nuclear energy, the implications this will have is that the probability of a person working in the nuclear field that at the same time opposes this technology is quite low.

Differences between genders are found in the majority of studies that investigate risk perception. Women express a higher risk perception for most risks we face, compared to men. From the available research, three models or strands can be identified to explain gender differences (see Davidson and Freudenberg 1996; Blocker and Eckberg 1997). These three models offer different explanations for understanding these differences and they are: differences in knowledge and familiarity with science including issues of trust; differences in risk

perception biological or social differences; and issues relating to cultural differences. It could be argued that differences due to biology or social processes should be treated as separate models. However, it has previously been argued that for the purpose of this chapter, whether these differences are socially constructed or in some manner appear “naturally” because of biological factors is not interesting. What is interesting however is how differences in risk perception are expressed and how this affects risk management in a larger perspective. If we were to acknowledge and accept that there are differences in perception between African, Asian, and Caucasian groups because of race and we would accept one groups’ perspective as superior to the others, we would (correctly) be accused of racism. Today no one would argue that one of those perspectives is better suited to deal with risk issues, but great care is taken to ensure that all perspectives are included in the risk debate. Many efforts have been taken, and some still remain to be realized, to ensure that different ethnic groups are not victims for an unfair balance of risk and benefit. To acknowledge and accept differences between men and women because of biology is equally sexist, still these claims are sometimes heard in the public debate to defend the position that one groups’ perception is better than the others.

In what follows, the three models will be further developed and empirical studies in these fields will be reviewed. Now, it is important to point out that none of these models claim to have the power to account for all differences found between men and women. Instead, they are to be viewed as partial explanations. They are also overlapping in many respects but will still be treated as three separate models. First in line is the model that used knowledge and familiarity with science as one explanation of the differences between men and women relating to risk perception.

## Knowledge of and Familiarity with Science

---

As risk issues are highly dependent on science, one way of explaining differences in perception of risks can be traced back to how much we know about the risks we face. Given this model, one explanation often used to understand women’s higher risk perception is related to women’s unfamiliarity with science and technology (Slovic 1999). Some scholars have suggested that opposition or skepticism toward new technologies is often interpreted as a lack of knowledge (Frewer 2004), suggesting that an increase of knowledge would alter risk perception. The “deficiency of knowledge”-assumption can also be used in order to discredit the opposition as they are portrayed as less knowledgeable and therefore not eligible to participate in risk discussions (Davidson and Freudenburg 1996).

If we examine the role of knowledge and risk perception, this would imply that increased knowledge of the scientific and technical aspects related to a risk will give a lower risk perception, whereas a lack of knowledge will make us more concerned. In most studies that focus on knowledge, men are more likely to state that they have more knowledge about technical environmental issues, or knowledge concerning risk sources in general as compared to women (O’Connor et al. 1999; Wester-Herber and Warg 2000). This reported greater knowledge leads men to have less concern for the environment, even though the link between knowledge and concern is not always clear (Blocker and Eckberg 1997). Indeed, this does not mean that women are less informed or men more knowledgeable – the opposite relationship can also be found. For example, those groups that feel most vulnerable to risk are the same ones that collect all information that is available to them on the issue that concerns them

(O'Connor et al. 1999). The link between an increase in knowledge and a reduced risk perception seems to have some empirical supports. However, a distinction needs to be made between reported knowledge and actual knowledge. It might also be worth asking that question exactly what is known: is it a better understanding of the probabilities, a wider grasp of the consequences, or a more sophisticated awareness balance between the risks and benefits?

In several studies, women are found to have a higher risk perception of the risks associated with GMOs, perceive fewer benefits, and have a lower acceptance of GMOs than men (Siegrist 2000; Kirk and McIntosh 2006; Simon 2011). It would seem then that women are less knowledgeable of the benefits with GMOs or have less knowledge about the probabilities of the possible negative consequences. The next logical step then would be to make women as knowledgeable as men in these issues. However, the differences in acceptance and benefits remain even though men and women have the same education (Kitto et al. 2003). In fact, it has been demonstrated that men and women with the same profession, and thereby the same education, judge risks differently (Slovic 1999). Another field where these differences are found is in the area of nuclear power or deposits of nuclear waste. In these studies, women generally express a higher risk perception than compared to well-educated white males (Greenberg et al. 2007; Greenberg 2009). It is not only in the complex technical risks that emerge from heavy industry that differences in risk perception between genders are found. In the new emerging information technologies, there are differences between men and women. In the context of social media, women are less likely to share personal information about them in social networks than men do, and also have greater concerns for privacy (Fogel and Nehmad 2009). Linked to this is the view of technical solutions to a range of risks. In the field of environmental perception, the view that technology can and will solve all ecological problems is more widely supported by men than women (Wehrmeyer and McNeil 2000). It should be stated, again, that these questions are rarely followed by a measurement of factual knowledge, implying that it is now known that men really *are* more knowledgeable – only that men more often state that they are.

## Issues of Trust

There are a few studies that look at the internal structure of the groups “men” and “women.” Some studies have added ethnic groups to the mix. Finucane et al. (2000) demonstrated that white males had a lower risk perception regarding both risks that individuals face and risks that the public are exposed to. This means that men are more likely to judge risks to them, for example risks of violent crime or natural disasters, as lower than what the same risk is to women or men of another ethnic group. In other words, there are studies that confirm that the groups of white males have a risk perception that is lower not only to that of women, but also other males of different ethnic groups. So when the level of education is the same, and the groups of white males are excluded, risk perceptions become more similar. Perhaps the reliance and confidence white males have toward technology is a reflection on the *faith* men put toward technology and technical solutions, a faith that other groups in society do not share.

Going back to the one study that diversified the group of men to include ethnic groups was found that white males expressed a higher confidence in technology managers and lower trust in government than the other groups (Finucane et al. 2000). This group of white males also reported to be not as sensitive to potential stigma that can be associated with high-risk ventures

as the other groups were. Also in other contexts, a difference in confidence toward governing authorities has been found. Regarding information security, men have been found to express lower confidence and trust in government than women do, even though both men and women express a view that personal information is at high risk of being misused (Wester and Sandin 2010). Men also tend to place higher trust and confidence in technology and technological progress than women do. This link is found in risk perception studies that measure the perception of man-made or technical risks (Siegrist 2000), but have not been fully explored in the environmental risk perception field (Blocker and Eckberg 1997). Along the same lines, as noted above: women often report that they intend to engage in more pro-environmental behavior, whereas men tend to place more trust and confidence into governmental policies that address environmental issues (O'Connor et al. 1999; Sundblad et al. 2007). However, as there is some evidence that even if women express more worry, they also report to have more hope in changing human behavior to make the world a better place; report to be more goal-oriented and more inclined to think about new and innovative ways of addressing environmental issues (Eisler et al. 2003). It is also found that women have stronger moral objections to technological developments where the consequences are seen as immoral or unethical. In fact, some studies suggest that men are more accepting of unethical actions than women (James and Hendrickson 2008). In similar contexts, studies suggest that men express a higher ideological motivation as one factor influencing their decisions to face risks compared to what women express (Billig 2006). This implies that values and ideology have a strong impact on perception and behavior, but in different ways from men and women. As will be developed in greater detail further on, women are more willing to act in a pro-environmental manner, even if the values toward the environment are similar between men and women. It could be wise then to focus on what options are available for action, and if these differ from men and women.

However, the differences due to ethnicity indicate that differences between genders cannot be explained by biological factors, nor can it be sufficient to relate to undiversified gender roles. Instead, differences are due in part to individual factors that are influenced by values, beliefs, and social roles. These roles are affected by both genders but also other aspects of group membership, where ethnic belonging is one. It would seem then that men tend to put more faith in technology and science than women do, and that men are more distrustful toward government. The interesting question then becomes why the group of (white) males have this perception: Are these differences due to nature or nurture? This will be developed below.

## Differences in Risk Perception due to Biological or Social Differences

The function of testosterone in the human body is to develop male sex characteristics, just like estrogen developed female sex characteristics (Seifert and Hoffnung 1997). These characteristics are both physical, such as the development of facial hair and muscle tissue, but are also believed to have an effect on the personal characteristics. Testosterone's influence on the male character and male behavior is seen primarily in relation to aggression, competitiveness, and the male libido. In many situations, it is found that men tend to engage in more risk-taking behavior than women. This is found in studies that report findings on different risk-taking activities such as sporting, such as skiing or mountain climbing (Lupton and Tulloch 2002); in risk-taking behavior in online social networks (Fogel and Nehmad 2009); and in financial

risk-taking (Barber and Odean 2001). Women on the other hand, report to be more afraid of nature, such as being afraid of spiders, lightning, or darkness, to a greater extent than men (van den Berg and ter Heijne 2005). Another possible explanation is that women experience emotions more vividly than do men, presumably due to biological factors. This would suggest that women react more emotionally to risks, rather than cognitively, and this would account for the higher risk perception of women (Sundblad et al. 2007).

Also, a majority of studies indicate that men are more risk-taking than women but that this difference decreases with age (Byrnes et al. 1999). It might be very tempting to interpret these differences as a willingness among men to be bold and daring, whereas women's actions are seen to be protective and cautious. Presumably also these differences would be caused by the male sex hormone: testosterone drives men to be daredevils and take risks as dictated by their biology. As men get older, and the levels of testosterone decrease, so does the frequency of risk-taking. However, looking at the statistical probabilities in some areas of having an accident or being injured, it reveals that women have a more realistic risk perception. For example, young male drivers systematically underestimate the risk of accidents even though this group is overrepresented in the statistics on motor vehicle accidents (Andersson and Lundborg 2007). Some studies report that men feel more fascinated by threatening situations, display less negative emotions in these situations, and would not avoid these situations in the future. However, it is difficult to provide empirical support that these differences are explained by differences in sensation-seeking (van den Berg and ter Heijne 2005). Put in other words, the subjective risk perception of males is lower than the objective risk and not because men are more prone, for one reason or another, to exposing themselves to dangerous situations. These results might encourage a view that looks at what groups are fearless, rather than those who are fearful (Kahan et al. 2007). These traits or characteristics are found elsewhere and there is no evidence that the lower risk perception and threshold for risk found among men is more objective or functional than the perception of women (O'Connor et al. 1999). Of course, there are situations where biological differences between men and women are relevant in risk research. For example, due to biological differences women can be more vulnerable to exposure of toxins, making it necessary to be more sensitive to gender in toxicological and epidemiological studies. In general, these studies fail to include gender differences caused both by biological and psychosocial differences. For example, in some cases there are differences in sensitivity to toxins caused by biological factors, such as women having more body fat than men due to higher levels of estrogen, but the (social) division of labor can also make women more exposed to toxins as they occupy other positions than men (Vahter et al. 2007). From this perspective, both biological and social differences between men and women need to be considered.

For a moment, we might consider that there are biological differences that cause the differences in perception of risk between men and women. Does the risk perception also affect behavior, and if so what does it mean for one of the biggest risks we face today? If men are driven by their biological wiring to take more risks, how will it affect their views on the environment and responses to the risks of environmental degradation? In many, but not all, studies on attitudes toward the environment, women show more concern for the environmental and hold more pro-environmental values than men. Women are found to have greater concerns about health and safety as they in their role are the primary caregivers to children and elderly. This, we could argue, can be caused by biological factors where women are different from men because of their childbearing abilities (Wehrmeyer and McNeil 2000).

One reason that women see higher risks than men can then be explained by, because of their biological functions, women take more responsibility for the family, such as being the primary caretaker for children and domestic chores. This would make women more observant toward risks that relate to their children (Lupton and Tulloch 2002). Included in this general framework is the explanation that differences between men and women in the perception of environmental concern relate to parenthood. The reasoning goes like this: It is hypothesized that when people become parents, women will have a nurturing role and be more concerned for the environment. This implies that nature is important for children and therefore also for women, but the reasons for this are not developed. Men, on the other hand, will be driven toward economic growth in order to care for his family. This will produce the result that men will be less concerned with environmental degradation and more focused on using the environment for economic growth (Davidson and Freudenberg 1996; Blocker and Eckberg 1997). Besides assuming that men are either ignorant or in denial of the risk we face today and that the environment will not sustain mankind at all if pushed to its limit due to their biology, closer scrutiny of empirical data reveals that this assumption does not hold. If the above stated reasoning is correct, studies would find that women that have children and that are removed from the workforce and stay at home should have the highest risk perception and would also be most inclined to engage in pro-environmental behavior. However, studies reveal that this group is less likely to engage in recycling and is less willing to pay for nature conservation. The empirical evidence also suggests that it is only women with children under the age of 6 that are more “green” in their perception of the environment: women and men with older or no children tend to not believe that humans are harming nature or engage in pro-environmental organizations (Blocker and Eckberg 1997). Women who stay at home on a full-time basis are also more prone to agree with statements that endorse salience of the economy over the environment, compared to women that work outside of the home. Perhaps this is because homemakers have to rely on the continuity or even growth of the economic capacity of their husbands in order to be able to stay at home and remain homemakers. This can make women dependent on men, but this is hardly due to biological characteristics.

These results suggest that parenthood seems to influence women to be more caring for the environment, but only for a limited time. Does this suggest that men are influenced by their hormones all the time, whereas the hormonal influence on women’s attitudes varies more during her life-course? Looking at biological differences between men and women might not be useful or constructive, as at best the male biology makes men underestimate risks whereas the female biology causes women to be more realistic, and these differences will fade over time. This is not the general view on how sex-specific hormones affect us as men and women. Perhaps it is more fruitful to examine the social roles that men and women inhabit and how this leads to specific differences in behavior. If men and women perceive risks differently, they will also respond to these risks in different ways – even if these behavioral differences have no biological origin or function.

The same explanation model – differences in attitudes toward economic growth and environmental risks – can be used, but instead of arguing that the differences between genders are due to biology, they can be seen as differences due to (social) gender roles. According to our gender-specific stereotypes, men are more often seen as the providers for their families and in that sense they might be more directed toward economic growth and more oriented to see benefits and be more accepting of risks. In this way, differences between men and women in economic growth orientation can be one explanation to differences in, for example,

environmental risk and concern for the environment. As men are more likely to believe that economic growth is important, men are also more inclined to have less concern for the environment. This argument builds on two assumptions: First, human beings in Western culture use the environment as a resource to further the growth of science and technology for economic progress. Second, through different social processes and through the division of labor, men are encouraged to engage in science and technology, whereas women are denied entrance to this field (Blocker and Eckberg 1997). Now, the discussion of how nature is perceived is outside the scope of this chapter. It might suffice to say that there are those that differences between men and women in how they perceive nature and what is believed to be natural. Instead it is argued here that the differences in environmental risk perception would be a result of how men and women occupy different roles in their lives and that these roles include behavioral responses that are limited by the options that are provided. An investigation of pro-environmental behavior reveals there is often an expressed behavioral intent among women that indicate that they would be more willing to engage in voluntary pro-environmental actions than men (Davidson and Freudenberg 1996; O'Connor et al. 1999; Slovic 1999; O'Connor et al. 1999). Women are also more likely to justify that it is never too expensive to regulate risks, implying that women are more likely to defend a higher cost in order to reduce risks (Slovic 1999). Interestingly enough, this tendency is higher among women even if men have similar environmental values and risk perception. However, issues of environmental concern and risk perception are complex and it is not possible to divide the population into clear and distinct groups, where some never engage in pro-environmental behavior and others do nothing but behave environmentally conscious in every situation. Also, that women show more concern for the environment does not imply that men are less willing to address environmental issues or that women are "eco-angels" (Wehrmeyer and McNeil 2000). Environmental degradation is an issue that concerns and engages many groups in society, not exclusively women. Individuals are complex and support some measures that aim to protect the environment and oppose others (O'Connor et al. 1999). However, the tolerance for and willingness to accept a certain degree of environmental damage seems to be more pronounced among men (Eisler et al. 2003).

Where does this leave us? If we summarize the two models that we have covered so far, it can be concluded that the effect of biology is limited but social roles and knowledge are factors that seem to matter. If we choose to focus on explanations that include social processes, this means that the differences found between men and women are reflections of the social context around us. Men are seen as the providers for the family, whereas the role of women is to pay closer attention to the needs of the family. This is also reflected in the division of labor in our society. Specifically, it means that men will occupy certain professions that will affect, or be affected, by their risk perception. In most areas that deal with progress of science and technology, where risks are produced, men more frequently occupy the role as experts than women do (Sjöberg 2002). It can then be argued that the risk perception of men is closer to the expert view of risk and that this perception is more often believed to reflect knowledge and science. However, a combination of these two models presented above suggests that the risk perception of women is in some cases more realistic than the perception men have, even if men claim to have more knowledge on the risks themselves. Certainly not all men can be experts, nor do all men have professions that allow them to influence the risk-producing or risk-governing processes that would make them less prone to experience a high-risk perception. So if the common link between different groups of men (and women too for that matter)

is not shared education or shared professional norms, what other factors can be present that link the gendered-role of men together? Perhaps examining the function of cultural groups can be one possible way.

## Cultural Explanations

---

Perhaps one of the most acknowledged theories that argue that men and women have different perceptions toward risks and particularly environmental risks is the ecofeminist theory. Theories found within ecofeminism are as diverse as the larger feminist research field (Longenecker 1997). Ecofeminist theorists, however, are similar in that they treat the differences between men and women in the perception and actions toward nature as closely linked to gender inequalities that are also found in other part of society. Parallels have been identified between the oppression of women and the exploitation of nature, making pro-environmental action more relevant to women rather than men (Agarwal 1992). Another more a critical position, where factors such as the dualism between nature/culture, male/female is explored to explain gender differences is also found in ecofeminist theories. Here it is argued that men are cast in the role as exploiters of the environment and women as the caretakers as a results of social processes, not biological ones (Agarwal 1992; Blocker and Eckberg 1997; Longenecker 1997; Wehrmeyer and McNeil 2000). This approach rhymes well with the explanation models presented above, where in the risk perception literature the same conclusion has been drawn. Now, ecofeminist theories have been criticized for treating women as a homogenous group, but attempts have been made to address this and bring about a more nuanced image of women (Agarwal 1992; Mellor 2007).

Another model that attempts to explain differences between men and women starts from the observed differences between white males and men and women of different ethnic groups found in the risk perception literature (Flynn et al. 1994; Finucane et al. 2000). The model of cultural groups focuses on the role of worldviews as they are expressed through a cultural theory of risk (Kahan et al. 2007). The cultural theory of risk suggests that there are differences in peoples' perception that, along with convictions on how society should be organized, makes them concerned with different risks. This model also suggests that the groups will prefer different risk management strategies. There are four dimensions, or "types," that individuals sort under: egalitarians, hierarchists, individualists, and communitarians (Sjöberg 2000; Kahan et al. 2007). Although this model has been criticized for being difficult to prove empirically (Sjöberg 1996, 2000), it has the advantage of offering an explanation that addresses the sociocultural differences between men and women in relation to risk perception. White males are found to hold more individualistic and hierarchical views than women or men of other ethnic groups (Finucane et al. 2000). This means that men are more likely to feel that they are in control over the risks to their health and that it is alright to impose small risks on people in order to receive a benefit. It has also been found that when an individual belongs to a cultural group, he or she is more likely to dismiss information that threatens the position or views held by the group (Cohen 2003). In other words, if white males believe that the world is a safe place; that risks are regulated properly and to a sufficient cost; believe that science and technology will solve ecological problems – it is highly unlikely that this perception will change. This cultural models presented here is very similar to the model presented previously, that differences in perception are due to social differences between men and women. Adding a cultural

perspective implies that these differences can be seen in a wider perspective. From this larger perspective, it can be seen that sociocultural norms and expectations are likely to cause differences in perception between men and women, rather than biology. Since the 1980s, the notion of gender differences, rather than sex differences, has made studies of men and women more sensitive to issues like power, influence, and control that are present both within families but also in larger social spheres (Ferree 1990). Adopting a gendered perspective implies that more attention is put toward identifying structures and roles, rather than focusing on biological differences. Women are *expected* to take a greater role in matters that relate to social life, which involves children, social relationships, and a consideration for the future of these relationships (Eisler et al. 2003). Men, on the other hand, are *expected* to be more risk-taking and involved with technical progress. In some ways and situations, women face a different reality than men. The division of labor and power inside the family to make formal decisions can lead to men making decisions that place women at risk.

## Cultural Divisions: Practical Implications

Before leaving the domain of cultural explanations for differences in perception, I would like to comment on how risks affect women differently from men, differences that are a result of social and cultural processes with a particular focus on the division of labor.

It can be argued that men and women face different realities or situations and this can be a contributing factor to differences in risk perception. One example of how the gender-based division of labor and how risks affect women can be found in the use of pesticide. In traditional farming in certain parts of the world, it is the men that decide how to conduct farming, what to grow, and what pesticides to use. The actual labor is carried out by women that work in the field. In combination with lower literacy rates and lower levels of education among women, this means that women are exposed to risks with pesticides but lack the formal training on how to handle them properly (Atreya 2007). It also means that the decision is outside of their domain, as women are not included in the decision-making process. In recent years, the frequency and extent of extreme weather phenomena has increased. This means that risks are moving from probabilities to actual events that need to be responded to. However, in preparing for disasters, gender is rarely taken into full consideration and the vulnerability or needs that are particular to women are often neglected (Norris et al. 2005; Enarson and Morrow 1998). In disaster response, women are greatly involved in relief and recovery efforts, but usually as part of the implementation not the decision-making or planning process (Noel 1998). This is done despite the recognition that women, as primary caretakers of their families, are vital in the disaster preparedness or response phase. In disaster response, it is found that women take warnings more seriously than men and also perceive to have the main responsibility for children. This is also evident after a disaster has struck, as women take on more responsibility for providing the immediate necessities. One example of the different situations facing men and women in the aftermath of a disaster is that men are free and encouraged to seek work outside of the community, where women remain in often unsecure environments to care for the children or elderly (Enarson and Morrow 1998). The same research team concluded that women are often in greater need of counseling and protection from abusers in the aftermath of a disaster, but it is doubtful that efforts of this kind is undertaken in a larger disaster response perspective. This lack of response that specifically

target women, and the areas over which women are responsible, is also present in how victims of disasters are portrayed. During the 9/11 disaster, men were more frequently depicted as victims in their roles as fathers, husbands, and sons and women portrayed as the grieving party (Monahan and Gregory 2001). The occupations as firefighters, police, and other first responders, have a long tradition of being predominately male, making women even more invisible from this arena. This connects back to the discussion above, where it is often found that men occupy certain professions, particularly those that involve science and technology. This means that women are mostly absent from fields where risks are created, identified, regulated, and responded to. Women report feeling more socially isolated during normal conditions and this becomes worse during a disaster as the social bonds are dispelled (Jackson and Henderson 1995; Norris et al. 2005). There is a danger however, to perceive women as vulnerable by default and to oversimplify the needs of women as a group (Fordham and Ketteridge 1998) but this does of course not mean that the special needs of women are to be ignored. Instead, this calls for an increase of gender-sensitive education that addresses the risk perception and response to them. However, even if increasing information and educational efforts to women in order to get a more “accurate” risk perception is acknowledged, it will not affect or alter the decision-making processes within the families. It will be interesting to see when observed differences between genders in risk perception also become a priority in risk research.

## Summary

---

It would seem that all models share one thing: differences in gender are due to differences in cultural roles. We learn early on what it means to be a man or a woman, and these roles are communicated through social and cultural expressions. These roles affect our values, where men are more likely to value individualism, favor economic growth, trust and rely on technical developments to make the world a better and safer place. Women on the other hand are more likely to value the environment, express more reservation toward technological development, and engage in sustainable behavior than men.

In sum, the models presented above would suggest the following:

The identification, assessment, and management of risks are highly dependent on science and technology. Some would also argue that risks stemming from the modern society are also created in this realm. What risk we choose to focus on, and perhaps to some extent also produce, is determined in large part due to what values we have. If economic development and growth is prioritized, other areas will not receive the same attention or resources. In our society, developments in the fields of science and technology are valued highly, making this field dominated by experts that have training in science and technology. It would seem then, that especially in areas of risk to health and the environment, the division between men and women becomes very clear. Usually these issues touch upon matters of scientific evidence, cutting-edge research, allocation of research funds (Jasanoff 1987), and the status that comes along with having influences on policy and future development of society (see, e.g., Oreskes and Conway 2010). This means that because of the way society is divided, in terms of education and profession, men are more likely to occupy the roles where risks are created and defined. Most of us can name few female scientists, save for Marie Curie, whereas the names of male

scientists are readily available from memory. This is not a coincidence, or collective laps in memory, but a reflection of what roles women have occupied within science (Watts 2007). Even if today the number of women involved in higher education and scientific work is increasing, the division of different social groups, of men and women, begins at early ages. Vocational choice is influenced not only by personal preference, but also by social processes, pressures, and norms. When it comes time to choose education that will eventually lead to a profession, the gap between girls and boys becomes evident. If we define some risks as a result technical progress, it can be worth noting that a technical university, such as my own, has about 30% female students and women make up about 10% of the senior faculty. Even though there are fields where women are increasing, such a biotechnology, this raise in women's presence often happens as the field has lost its initial high status and is becoming more mainstream (Rosen 1994). It might be worth reflecting over how institutes that are developed to ensure that the environmental risks that we face, such as the Intergovernmental Panel on Climate Change, has predominantly men – albeit not exclusively white men – in the highest positions whereas women are found on lower positions as cochairs and support staff.

It is tempting then, in order to address this issue is to increase the number of women in science and technology but this is not my point. If we strive toward making women more involved in these spheres, we run the risk of heading down a dead-end road (or at least a cul-de-sac). First, involving more women in science and technology in order to make women more educated on these matters, tends to support the idea that if only the level of knowledge was equal, so would the risk perception between men and women be. In this way, we make the same mistake as mentioned above and disregard that risk issues are issues where values, ideologies, and power are equally important as any scientific calculation. Second, if we include women not to increase their knowledge but to increase the female representation in these spheres, what do we expect women to contribute with? Do we expect women to provide a softer perspective on hard technological issues? Do we assume that women will add a gentler touch and be more considerate to vulnerable groups in society? Whatever contribution we assume that women will make if only their number were increased in the risk debate, we need to be careful and rethink what that contribution might be. If not, we apply the same stereotypical models and expectations of what men and women *are expected* to contribute and the time has come to move beyond these stereotypes.

## Further Research

It was stated in the beginning of this chapter that the risks presented here were risks that to a great extent lay outside of an individual's personal control and risks that are impossible to escape from. Despite this, men and women perceive these risks differently even if we can agree that the consequences are the same for all groups if things do wrong. If we live next to a high-risk facility, we will be equally affected from a physiological perspective, even if we *perceive* the risk differently. In light of this, we need to pay more attention to why there are differences between men and women in risk perception and especially how these differences express themselves in risk management. It was concluded above that knowledge and familiarity are important for risk perception. One suggestion for future research can be to critically examine these claims of greater knowledge that men have, but not in a general sense. Measuring the level

of knowledge among the general public on risk issues does not seem to bring this research field forward. Instead, examining men and women with the same education that presumably have the same level of knowledge can help shed light over the role knowledge has in risk matters. Judging from the above discussion, I would like to put forth the hypothesis that knowledge plays an inferior role in risk perception when compared to underlying values and cultural belonging (or ideology). If future research can identify the relationship between estimated knowledge and “real” knowledge, perhaps we can move beyond the simple model of a knowledge-deficit among women.

The second research challenge calls for a critical examination of the function of stereotypes in risk matters. Examining the difference between men and women is not a simple question of discrimination. Forcing more women into technical fields or passing legislation that would require men to become homemakers or become more involved in the care for children and the elderly does not seem to be an appropriate response to the question of gender differences in risk perception. No one benefits from being cast as a victim, and in this chapter it has not been my intention to portray women as victims. Personal preferences and choices need to be respected, and living in a democratic society we are free to plot the course of our lives. Still the question that would need closer examination is how stereotypes limit life for all citizens, not just one group. Perhaps men suffer equally from the gender division in education and labor, or have difficulties finding options that enable them to act in manners that are more in line with their character and not with their gender role. Here, issues of power are important as they are central to risk debates. The second area I would like to suggest for future research is an increased attention paid to the inequalities that not only exclude women from risk discussions, but to focus on issues of power and the distribution of risk and benefit. Being able to identify and demonstrate how privileged groups, regardless of gender, influence decisions on risks is vital to a democratic society. My second hypothesis is that the internal variation within the group of men will be greater than the differences between men and women.

The third and final research direction I would like to suggest it by examining hard facts. Differences between men and women are present in many areas that involve risks: risks are perceived differently; risks are handled differently, but also the risks men and women face are different (Gustafson 1998). Women face different consequences from the same risk, than men do. For example, a woman walking home late at night might consider the probability of being attacked, robbed, beaten, and raped. For a man, that same situation can cause him to think about the probability of being robbed and beaten but perhaps not raped. This often makes women more concerned about risks of violence than compared to men (Lupton and Tulloch 2002). However, there is some evidence that white men that are wealthy are more fearful of being victims of a crime than men of more modest means, as the wealthy men have more to lose in terms of property and assets (Franklin and Franklin 2009). Often, this leads to two different judgments that are caused by the differences in consequences and not by the statistical probability. Another way of expressing this, rather than in terms of perceived risk, is perceived insecurity. Also in this area, studies find that women score higher than men, suggesting that women see the physical environment as more risky and less secure than men (Carro et al. 2010). This is also found in situations where individuals live under stressful conditions, such as in areas of conflict or violence, where men perceive the area as less risky than women (Billig 2006; Rodionova et al. 2009). The third suggestion I would make then is to document and demonstrate how men and women’s lives are affected by risk and crises on a very concrete level, and to let this empirical evidence have a real influence on the risk management process. I would encourage the

reader to closely examine the daily newspapers and television broadcasts during a crisis and search for images of women that are not portrayed as victims. My third hypothesis is as follows: there will be no images of women as first responders or any reporting of the special needs of women. This does not mean that there are none.

## References

- Agarwal B (1992) The gender and environment debate: lessons from India. *Fem Stud* 18(1):119–158
- Andersson H, Lundborg P (2007) Perception of own death risk. *J Risk Uncertain* 34:67–84
- Atreya K (2007) Pesticide use knowledge and practices: a gender differences in Nepal. *Environ Res* 104:305–311
- Barber BM, Odean T (2001) Boys will be boys: gender, overconfidence, and common stock investment. *Q J Econ* 116:261–292
- Beck U (1986) The risk society: towards a new modernity. Sage, London/Newbury Park
- Bem SL (1983) Gender schema theory and its implications for child development: raising gender-aschematic children in a gender-schematic society. *Signs* 8:598–616
- Billig M (2006) Is my home my castle? Place attachment, risk perception, and religious faith. *Environ Behav* 38:248
- Blocker TJ, Eckberg DL (1997) Gender and environmentalism: results from the 1993 general social survey. *Soc Sci Q* 78:841–858
- Byrnes JP, Miller DC, Schafer WD (1999) Gender differences in risk taking: a meta-analysis. *Psychol Bull* 125:367
- Carro D, Valera S, Vidal T (2010) Perceived insecurity in the public space: personal, social and environmental variables. *Qual Quant* 44:303–314
- Cohen GL (2003) Party over policy: the dominating impact of group influence on political beliefs. *J Pers Soc Psychol* 85:808–822
- Davidson D, Freudenburg W (1996) Gender and environmental risk concerns. *Environ Behav* 28:302–339
- Deeks A, Lombard C, Michelmore J, Teede H (2009) The effects of gender and age on health related behaviors. *BMC Public Health* 9:213
- DeJoy DM (1992) An examination of gender differences in traffic accident risk perception. *Accid Anal Prev* 24:237–246
- Eisler AD, Eisler H, Yoshida M (2003) Perception of human ecology: cross-cultural and gender comparisons. *J Environ Psychol* 23:89–101
- Enarson E, Morrow BH (1998) The gendered terrain of disaster. Praeger, Westport
- Ferree MM (1990) Beyond separate spheres: Feminism and family research. *J Marriage Fam* 52:866–884
- Finucane ML, Slovic P, Mertz CK, Flynn J, Satterfield TA (2000) Gender, race, and perceived risk: the “white male” effect. *Health Risk Soc* 2:159–172
- Flynn J, Slovic P, Mertz CK (1994) Gender, race, and perception of environmental health risks. *Risk Anal* 14:1101–1108
- Fogel J, Nehmad E (2009) Internet social network communities: risk taking, trust, and privacy concerns. *Comput Hum Behav* 25:153–160
- Fordham M, Ketteridge AM (1998) Men must work and women must weep: examining gender stereotypes in disasters. In: Enarson E, Morrow BH (eds) The gender terrain of disasters. Praeger, Westport, pp 81–94
- Franklin CA, Franklin TW (2009) Predicting fear of crime. *Fem Criminol* 4:83
- Frewer L (2004) The public and effective risk communication. *Toxicol Lett*, 149(1–3):391–397
- Greenberg M (2009) Energy sources, public policy, and public preferences: Analysis of US national and site-specific data. *Energy Policy* 37:3242–3249
- Greenberg M, Lowrie K, Burger J, Powers C, Gochfeld M et al (2007) Nuclear waste and public worries: public perceptions of the United States' major nuclear weapons legacy sites. *Hum Ecol Rev* 14:1–12
- Gustafson PE (1998) Gender differences in risk perception: theoretical and methodological perspectives. *Risk Anal* 18:805–811
- Horowitz M (1976) Aristotle and woman. *J Hist Biol* 9:183–213
- Jackson EL, Henderson K (1995) Gender-based analysis of leisure constraints. *Leisure Sci* 17:31–51
- James HS Jr, Hendrickson MK (2008) Perceived economic pressures and farmer ethics. *Agric Econ* 38:349–361
- Jasanoff SS (1987) Contested boundaries in policy-relevant science. *Soc Stud Sci* 17:195–230
- Kahan DM, Braman D, Gastil J, Slovic P, Mertz CK (2007) Culture and identity-protective cognition: explaining the white-male effect in risk perception. *J Empir Leg Stud* 4(3):465–505
- Keesing RM (1981) Cultural anthropology: a contemporary perspective, 2nd edn. Holt, Rinehart and Winston, New York

- Kirk DD, McIntosh K (2006) Indications from a public survey. *AgBioForum* 8(4):228–234
- Kitto SL, Griffiths LG, Pesek JD (2003) A long-term study of knowledge, risk, and ethics for students enrolled in an introductory biotechnology course, 2. *J Anim Sci* 81:1348
- Kulick D (1987) *Från kön till genus* [From sex to gender]. Carlsson, Stockholm
- Longenecker M (1997) Women, ecology, and the environment: an introduction. *NWSA J* 9:1–17
- Lupton D, Tulloch J (2002) Risk is part of your life? risk epistemologies among a group of Australians. *Sociology* 36:317
- Mellor M (2007) Ecofeminism: linking gender and ecology. In: Pretty JN et al (eds) *The SAGE handbook of environment and society*. Sage, London, p 66
- Merchant C (1980) *The death of nature: women, ecology, and the scientific revolution*. Harper & Row, San Francisco
- Monahan B, Gregory C (2001) From ground zero to ground hero: status appropriation and the FDNY. Disaster Research Center, Preliminary paper 315
- Noel GE (1998) The role of women in health-related aspects of emergency management: A Caribbean perspective. In: Enarson E, Morrow BH (eds) *The gendered terrain of disaster: through women's eyes*. Praeger, Westport, pp 213–223
- Norris FH, Baker CK, Murphy AD, Kaniasty K (2005) Social support mobilization and deterioration after Mexico's 1999 flood: effects of context, gender, and time. *Am J Community Psychol* 36:15–28
- O'Connor RE, Bord RJ, Fisher A (1999) Risk perceptions, general environmental beliefs, and willingness to address climate change. *Risk Anal* 19:461–471
- Oreskes N, Conway EM (2010) *Merchants of doubt: how a handful of scientists obscured the truth on issues from tobacco smoke to global warming*. Bloomsbury, New York
- Rodionova N, Vinsonneau G, Riviere S, Mullet E (2009) Societal risk perception in present day Russia. *Hum Ecol Risk Assess Int J* 15:388–400
- Rosen A (1994) Adam, Eve and the controversial rib: gender, technology, conflict and universalism. *Dialogue Hum* 4:23–30
- Seifert KL, Hoffnung RJ (1997) *Child and adolescent development*, 4th edn. Houghton Mifflin, Boston
- Siegrist M (2000) The influence of trust and perceptions of risks and benefits on the acceptance of gene technology. *Risk Anal* 20:195–204
- Simon RM (2011) Gendered contexts: masculinity, knowledge, and attitudes toward biotechnology. *Public Underst Sci* 20(3):334–346
- Sjöberg L (1996) A discussion of the limitations of the psychometric and cultural theory approaches to risk perception. *Radiat Prot Dosim* 68:219–225
- Sjöberg L (2000) Factors in risk perception. *Risk Anal* 20:1–12
- Sjöberg L (2002) The allegedly simple structure of experts' risk perception: an urban legend in risk research. *Sci Technol Hum Values* 27:443
- Slovic P (1999) Trust, emotion, sex, politics, and science: surveying the risk-assessment battlefield. *Risk Anal* 19:689–701
- Sundblad EL, Biel A, Gärling T (2007) Cognitive and affective risk judgements related to climate change. *J Environ Psychol* 27:97–106
- Vahter M, Gochfeld M, Casati B, Thiruchelvam M, Falk Filippson A, Kavlock R, Marafante E, Cory-Slechta D (2007) Implications of gender differences for human health risk assessment and toxicology. *Environ Res* 104:70–84
- van den Berg AE, ter Heijne M (2005) Fear versus fascination: an exploration of emotional responses to natural threats. *J Environ Psychol* 25:261–272
- Watts R (2007) *Women in science: a social and cultural history*. Routledge, London/New York
- Wehrmeyer W, McNeil M (2000) Activists, pragmatists, technophiles and tree-huggers? Gender differences in employees' environmental attitudes. *J Bus Ethics* 28:211–222
- Weintraub S et al (1984) The development of sex role stereotypes in the third year: relationships to gender labeling, gender identity, sex-types toy preference, and family characteristics. *Child Dev* 55: 1493–1503
- Weisaeth L, Knudsen Ø, Tønnessen A (2002) Technological disasters, crisis management and leadership stress. *J Hazard Mater* 93:33–45
- Wester M (2009) Cause and consequences of crises: how perception can influence communication. *J Contingencies Crisis Manag* 17:118–125
- Wester M, Sandin P (2010) Privacy and the public – perception and acceptance of various applications of ICT. In: Arias-Olivia M, Ward-Bynum T, Rogerson S, Torres-Coronado T (eds) *The "backwards, forwards and sideways" changes of ICT*, 11th international conference on the social and ethical impacts of information and communication technology (ETHICOMP), Tarragona, pp 580–586
- Wester-Herber M, Warg L-E (2000) Gender and regional differences in risk perception: results from implementing the Seveso II Directive in Sweden. *J Risk Res* 5:69–81

# 42 Risk and Soft Impacts

Tsjalling Swierstra<sup>1</sup> · Hedwig te Molder<sup>2</sup>

<sup>1</sup>University of Maastricht, Maastricht, The Netherlands

<sup>2</sup>University of Twente/Wageningen University, Enschede/Wageningen,  
The Netherlands

<i>Introduction</i> .....	1050
<i>History</i> .....	1050
<i>Current Research</i> .....	1051
<i>It Is Not All About Health: How Soft Concerns Tend to Get Overlooked</i> .....	1053
<i>How Soft Impacts Tend to Disappear from the Public Agenda: The Case of “Naturalness”</i> .....	1056
<i>Three Dimensions of the Hard/Soft Distinction: An Explanatory Model</i> .....	1058
Valuation .....	1058
Quantifiability .....	1060
Causality .....	1061
<i>Concluding Remarks</i> .....	1062
<i>Further Research</i> .....	1063
<i>Notes</i> .....	1065

**Abstract:** Policy and technology actors seem to focus “naturally” on risk rather than on technology’s social and ethical impacts that typically constitute an important focus of concern for philosophers of technology, as well as for the broader public. There is nothing natural about this bias. It is the result of the way discourses on technology and policy are structured in technological, liberal, pluralistic societies. Risks qualify as “hard” (i.e., objective, rational, neutral, factual), other impacts as “soft” (i.e., subjective, emotional, partisan, value-laden) and are therefore dismissable. To help redress this bias, it is necessary to understand how this distinction between hard and soft impacts is construed – in practice and in theory. How are expected (desired, feared) impacts of technology played out in expert-citizen/consumer interactions? We first discuss online patient deliberations on a future pill for celiac disease (“gluten intolerance”) promising to replace patients’ lifelong diet. By “rejecting” this pill, patients displayed concerns about how the new technology would affect their identity, and the values incorporated in the way they had learned to handle their disease. Secondly, we analyze how experts construct a consumers’ concern with “naturalness” of food: as a private – and invalid – preference that requires no further debate. The point of the analysis is to make available for discussion and reflection currently dominant ways to demarcate public and private issues in relation to emerging technologies, including the accompanying distributions of tasks and responsibilities over experts and laypersons. However, the actors themselves cannot simply alter these demarcations and distributions at will. Their manoeuvring room is co-shaped by discursive structures at work in modern, technological, pluralist, liberal societies. In the third section, we therefore identify these structures, as they provide the hegemonic answers to the three key questions with regard to the possible impacts of emerging technologies: how are impacts *evaluated*; how are they *estimated*; and how are they *caused*? We conclude with some suggestions for further research.

## Introduction

---

“Risks” typically concern harms to values like health, environment, and safety. But the larger public sometimes is interested in another type of consequences of existing or emerging technologies as well, positive or negative consequences that we refer to as “soft impacts.” Until now, these soft impacts receive relatively little attention in Risk Studies, and go largely ignored by policy makers and technologists. In this chapter we show how concerns with soft impacts often get overlooked. And if they are acknowledged, they typically get subtly removed from the agenda. We offer some explanations for this exclusion of soft impacts (Swierstra et al. 2009; Boenink et al. 2010), and conclude with some suggestions for further research.

## History

---

For a long time scientific and technological progress seemed to equal societal progress. From the 1950s onward, however, the conclusion became inescapable for policy makers and technology actors that technological innovations can and often do have unintended, unforeseen, and/or undesirable impacts. Risk assessment was invented to warn society in advance for such impacts, and thus help to avert them by taking social and/or technical precautions.

It is interesting to observe that the growing awareness of technology's unintended and unwanted impacts during the previous decades was hardly informed by the philosophy of technology. Classic philosophers of technology, e.g., Martin Heidegger and Jacques Ellul, had already devoted ample attention to technology's darker side at the time when "technological risk" became prominent on society's agenda. But they tended to focus less on safety, health, or environmental issues, concentrating instead on the consequences of new and emerging technologies for

- Established meanings, world and life views (cultural)
- Existing values, norms, and conceptions of the good life (moral)
- The (global) distribution of power and control (political)

Some complained about technology because it eroded tradition, replacing it by uniformity and conformism. They feared dehumanization, depersonalization, spiritual shallowness, desensitizing, and mind-numbing as a result of automation, and in the end, the substitution of humans by machines. Others stressed technology's moral consequences: technology would lead to the devaluation of life's fundamental values, cause moral corruption, and result in eternal unhappiness or shallowness, through the creation of artificial needs. A specific variety of the moral corruption thesis is couched in religious terms: technology was accused of creating false gods and of giving the false illusion that man is no longer dependent on God, thus leading man to commit the sin of hubris. Again others warned that technology would help create new tyrannies, that would be all the more secure because of psychological manipulation. These tyrannies would undermine our privacy through observation techniques and data banking, or their anonymous systemic logic would marginalize democratic deliberation.

Of course, these doom prophets invited all kinds of reassuring rebuttals by other philosophers, who argued that this ink-black pessimism was ungrounded. In fact technology had exactly the opposite impacts: enriching culture, strengthening morality and religion, and enhancing democracy (For an overview, see Van der Pot 1985).

Thus far, policy makers and technologists by and large ignore these discussions. In the past that could be justified by the fact that according to many philosophers technology was inseparable from its unwanted consequences. Because of this technophobic bias, their work held little promise for policy makers and technologists who were faced with the practical task to make technology safer, but were not prepared to throw out technology altogether. But this situation has changed drastically since. Most modern philosophers of technology are no longer in the business of dismissing Technology (with a capital T). Since the "empirical turn" (Achterhuis 2001) they tend to study the impacts of specific technologies in specific contexts, without a priori leaning toward pessimism of optimism.

## Current Research

Like their predecessors, modern philosophers of technology still tend to focus on a *different* type of impacts than is common in risk studies. In this chapter, we will argue that it is important to broaden the assessment of technology's impacts from risk to the kind of "soft impacts" that are typically in the center of attention of philosophers of technology. We offer two reasons for this broadening of the agenda. The first one is that many laypersons worry about these soft impacts, and therefore democracy requires that at least they are being assessed

and discussed openly. The second reason is that Technology Assessment aims at better technology. To realize this aim, it is important to take a wide array of possible impacts into account, not only risk.

However, getting soft impacts on the agenda of policy makers and technologists is not a simple matter. Broader cultural, moral, and political aspects are regularly voiced in public discussions, but seem to have difficulty gaining access to the agendas of policy and technology actors. These parties “naturally” seem to focus on risk rather than on technology’s social and ethical impacts. Or rather, their focus is on risk assessment and everything else is dubbed an “ethical issue.” This framing then makes the prevalent “non-risk” issues ready to be recognized as legitimate but solely private concerns (Wynne 2001; Swierstra 2002), which are out of place on the public agenda.

We will argue that there is nothing natural about this bias. It is the result of the way discourses on technology and policy are structured in technological, liberal, pluralistic societies. Impacts of (emerging) technologies that qualify as “hard” (i.e., objective, rational, neutral, factual) attract much more attention than impacts that can be dismissed as “soft” (that is, subjective, emotional, partisan, value-laden). And risks qualify as “hard,” social and ethical impacts as “soft.” While the relation between soft impacts and the evolution of public controversy is not linear and direct, experience and research (see for example Marris 2001) have shown that the dismissal of latent concerns about soft impacts (soft concerns) may easily engender unexpected – at least for technologist designers – outbursts of public discontent later in time. By then, repeated experiences and cumulated irritations have replaced the early, largely invisible and not necessarily *negative* concerns. Technologists may feel nothing but annoyance about the public’s irrational moves – no longer being able to recognize that, for instance, religious critiques (“playing God”) might also pose questions about the limits of science (Wynne 2001). The paradigm case here is the Monsanto debacle of the mid-1990s. A lot of public concerns seemed to regard the *hard* impacts of modified crops – environmental risks and health concerns – but these concerns often sprang up from other concerns about *soft* impacts, e.g., that genetic modification exemplified technological hubris, or that it increased the power of big corporations over small farmers (Marris 2001).

To help redress this bias, we need to analyze how this distinction between hard and soft impacts is construed – in practice and in theory. As we are aware of, our concern with the impact of “other than risk issues” on public dialogue is not entirely new. Other studies, mainly in the area of science and technology studies (e.g., Jasianoff 2003; Hobson-West 2007; Wynne 1996, 2001), have pointed to the importance of seeking to evaluate technology’s aims rather than its mere consequences in terms of risk (Jasanoff 2003, p. 224), and “the uncritical framing of contemporary controversies as primarily about risk, or even about different understanding of risk” (Hobson-West 2007, p. 211). Brian Wynne’s work (e.g., 1996, 2001, 2006) perhaps most prominently refers to the significance of addressing wider social and political questions in public debate on emerging technologies.

While these authors do recognize the importance and dismissal of other than risk issues, little is known about how these demarcations between hard and soft impacts of technology are performed in real-life situations, for what purposes (consciously or not), and with what consequences. Furthermore, the question remains what these impacts, and the difference between them, actually consist of. (How) can they be characterized, and what makes them susceptible for more or less devoted attention? In this chapter, we will make a start with both questions.

First, we will analyze in close detail how expected (desired, feared) impacts of technology are played out in expert-citizen/consumer interactions. We first discuss an illustrative example of how soft impacts surface in online patient deliberations on an emerging technology, namely, a future pill for celiac disease (“gluten intolerance”) patients that was promised to replace their lifelong gluten-free diet. We show how these patients, by “rejecting” the proposed technology, displayed concerns about how the new technology would affect their identity, and the values incorporated in the way they had learned to handle their disease. Their rejection was targeted not so much at the pill itself but at the experts’ construction of their current life as highly problematic and the pill as a perfect solution for that problem. The example illustrates the indirect way in which soft concerns often manifest themselves.

In our second example, we study closely how subtly – respectfully – these soft concerns often get dismissed. To illustrate this point, we look at an example of expert interaction in which a consumers’ concern with the “naturalness” of food is both constructed as a private issue and discounted as nonvalid. We are claiming neither that the expert is wrong nor that the consumer or patient is right. The point of the analysis is to make available for discussion and reflection currently dominant ways to demarcate public and private issues in relation to emerging technologies, including the accompanying distributions of tasks and responsibilities over experts and laypersons.

However, it would be naïve to assume that the actors themselves could simply alter these demarcations and distributions at will. Their manoeuvring room is co-shaped by discursive structures at work in modern, technological, pluralist, and liberal societies. In the third section we therefore will identify these structures, as they provide the hegemonic answers to the three key questions with regard to the possible impacts of emerging technologies: how are impacts *evaluated*; how are they *estimated*; and how are they *caused*? Together these answers help construct some positions as being rational, public, neutral, and serious, and others as being irrational, private, partisan, and not to be respected. We conclude by pointing out why, if one aims for an open and comprehensive public dialogue about science and technology, it is crucial to modify these discursive structures.

## **It Is Not All About Health: How Soft Concerns Tend to Get Overlooked**

---

The point of medical technology is to help increase (or defend) our health. Therefore, it seems a pretty straightforward matter that discussions about emerging medical technologies would concentrate on these would impact our health. In reality, however, matters are not so simple. We will illustrate this by drawing on examples from a broader study of celiac patients’ accounts regarding a future pill (te Molder et al. [submitted](#); Veen et al. [2010](#)).

In our analysis, we applied a discursive psychological approach that starts from the assumption that talk is oriented to *action* rather than merely *reflecting* reality. So instead of determining the truth-value of what people report – by looking at what a person really wants, thinks, or feels, or what the world really looks like – the focus is on what people’s utterances *do* in the interaction, such as accusing, complaining, and complimenting (Edwards [1997](#); Potter [1996](#); te Molder and Potter [2005](#)). People use the turn-by-turn development of a conversation as a resource to make sense of each other’s talk. They may treat displays of anger as a request to

leave the room, claimed losses of memory as reluctance to answer a question, or deal with a description of their behavior as implicating blame. These continuously updated understandings of what is being said and done constitute an important “proof procedure” for the analyst, that is, he or she can use these displays to provide support for the analysis. Whether something is blame or compliment is not decided upon by the analyst in the first place but analyzed as a participants’ concern.

People also talk rhetorically, in that they routinely resist or deny actual or potential alternative versions of what is being said. Inspecting stretches of discourse for these alternative versions helps the analyst to make sense of the actions performed. Presenting yourself as a woman resists “being a man,” and that may provide cues for the action at stake, for example, in the context of alleged or claimed transsexualism. It is the combination of a sequential and a rhetorical analysis which forms the basis of a discursive psychological approach (te Molder 2008).

The discussion on the gluten pill is part of an online forum for celiac disease patients ([www.celiac.com](http://www.celiac.com)). Celiac disease is a genetic disorder that causes an autoimmune reaction to the wheat protein gluten, which results in serious damage to the small intestine. At the moment, a lifelong diet is the only remedy. This requires not only discipline but is also difficult to implement as gluten is found in many daily foods.

Now let us have a look at extract 1, in which a (self-reported) scientific expert introduces the pill. The focus is on what the expert’s question is *doing* – in terms of discursive action – by looking at how the participants at the online discussion forum *treat* his or her contribution:

### Extract 1

- 1 Researcher (Sept 6 2004, 09:38 AM)  
2 Newbie  
3  
4 I am doing some research on developing potential new therapies for celiac  
5 disease and am wondering, how much would you be willing to pay each day if  
6 you could take a pill that would let you eat a normal diet? How much would  
7 you pay per year?  
8 ((9 lines omitted))  
9  
10 Sammy (Sept 9 2004, 08:04 PM)  
11 Member  
12  
13 I wouldn’t give one red cent for a pill. I have taken pills all of my life  
14 because of this disease. I would just keep on with the diet as is. I feel  
15 better than ever and have more energy than most 60 year olds should have.  
16 Pills? Thanks any way. Sammy

The topic is initiated by a researcher, obviously not a celiac patient and in this respect an outsider on the forum. Notice how by requiring into the amount of money that patients would be willing to pay each day (lines 5–7), the issue of need or desire to have this pill is already answered for. Second, the pill is presented as an *easy solution* to the disease in comparison to the current treatment (“a pill that would let you eat a normal diet,” line 6).

Sammy's contribution challenges the validity of both presuppositions. By saying that she "wouldn't give one red cent for a pill" (line 13), she explicitly brings down the assumption that celiac patients would take the pill anyway. She grounds her rejection in her elaborate experience with pills (lines 13–14). If you have used pills all your life, and the disease has ultimately been treated effectively by a diet, it makes no sense to go back on a pill and give up the diet and its payoff ("better than ever" and "more energy than..." line 15). Sammy's reply thereby questions the assumption in the researcher's post that the pill will radically change her life for the better.

Interestingly, the question including the presupposition that celiac patients will take and need a pill no matter what, evoked much stronger reactions than the careful suggestion that a pill might be developed:

**Extract 2** If they came out with an anti-gluten pill thingy (IV, 1–2; 4; 6–7)

- 132 If they found a pill that would neutralize the effects of  
gluten on your body (sort of like the pill people  
133 take who are lactose intolerant), would you use it?  
134  
135 Yes, definitely – all the time [18] [43.90%]  
136 Sometimes, but only when I am eating out [12] [29.27%]  
137 Sometimes, maybe once or twice a week [4] [9.76%]  
138 No, I'd be afraid that it wouldn't work [4] [9.76%]  
139 No, I don't think I could ever look at wheat the same way [3] [7.32%]  
140 Total Votes: 41  
141  
142  
143 Ronald (Apr 9 2004, 12:35 PM)  
144 Advanced Member  
145  
146 It could happen, eventually.....

This then shows that it is not the pill *itself* which is disputable, but the assumption that patients will use it *as a matter of course*. It is at this point that we become aware of the presence of concerns that do not regard health or safety ("risk") issues. Sammy, for example, rejects being characterized as a passive patient. She presents herself as a healthy individual who is able to maintain her vitality in the face of adverse circumstances. By resisting the notion that they would straightforwardly accept the pill patients construct themselves as proactive, thoughtful people with a healthy way of life. Presenting new possibilities as cure-alls makes the gluten-free diet appear as a hardship, and undermines the complexity of the patients' relation to their disease, including the positive values embedded in that relation. This example shows that an apparent straightforward rejection of a new medical technology is drawn upon by patients not so much to show concern about the pill's impact on their health, but about how the presentation of this innovation impacts their identity and sense of achievement (see also Veen et al. 2010; te Molder et al. submitted).

So, "less tangible" concerns often *emerge from* rather than stand out in discussions about new technologies. We find these, in this case, identity- and lifestyle-related concerns (who am

I – a patient, a victim, a naïve believer in cure, a healthy person?) only by looking at the ways in which patients *treat* the expert's contribution, and not so much by focusing on the content of what they say (e.g., I do not want the pill). This shows that these concerns are often only available indirectly for the analyst or debate facilitator. Moreover, participants themselves often do not have direct access to such interactional concerns. More precisely, we should say that this type of concerns regarding emerging technologies typically seem to arise as interactional goals – consciously or not – of what people say, rather than that they can be found directly in the content of the arguments that are put forward.

Now we turn our attention to an example that illustrates how experts operate to allow some concerns about technology's possible impacts access to the public agenda, while denying a similar access to other concerns. Again, this is typically done in a way that is far from straightforward.

## How Soft Impacts Tend to Disappear from the Public Agenda: The Case of "Naturalness"

The next fragment is part of a larger study of expert talk on future foods. It illustrates how a relatively classic citizen theme – naturalness, in this case of food – may be removed from the public agenda. In contrast to the previous example, in which an identity concern emerged from the discussion in such a way that it was neither available for experts nor patient participants, here "naturalness" appears as an explicit theme on the agenda. This can partly be explained by the fact that this discussion about the future of food was organized (not spontaneous, as in the first example) and the theme was put forward by the discussion leader. But naturalness is also a classic theme when it comes to citizen concerns about all sorts of new technologies. The argument is both attributed to citizens by experts and drawn upon by citizens themselves (e.g., Marris 2001 for naturalness in relation to food). It is treated as a typical citizen concern that is readily available and needs no further explanation, as we will also clarify with the following example.

The extract is taken from a discussion among twenty Dutch stakeholders about future food technologies, nine of whom were scientific and industrial food experts (Middendorp et al. *in prep.*). It illustrates how "naturalness" is removed from the public agenda by attributing the theme to the private domain of consumers such that no special account need be given, and no further exploration of its meaning is required:

### Extract 3<sup>1</sup>

- Facilitator 1 bu- but the picture that emerges now  
2 is of uh as it were  
3 an uhh (0.4) somewhat  
4 powerless industry  
5 that have to dance to the contradictory whims  
6 of the consumer (0.6)  
7 uhh is that the current feeling  
8 or are there also ideas about naturalness  
9 with the industry itself  
10 ((expert gets his turn from facilitator))

Expert      11 yes I think the industry  
12 views it a little bit-  
13 a little bit differently (0.4)  
14 uhh there are indeed (0.8) consumers  
15 who indeed want natural  
16 →without probably many consumers  
17 →uhh understanding what that then means  
18 →and what it entails (1.3)  
19 ehh subsequently (0.6) one wants e-number free  
20 well the industry can make it (0.7)  
21 the only problem of course is  
22 if you want to produce it e-number free  
23 that is more difficult that is more expensive  
24 the quality is generally less  
25 and it ultimately costs a little bit more (0.7)  
26 well if the consu- if the consumer wants that  
27 then I think that the industry simply has to  
28 ↑make it (1.1)  
29 as simple as that

We are interested in the kind of *action* that the expert performs by responding in the way he does. First note how the expert's remark about industry and consumers having different views (lines 11–13) avoids answering the facilitator's question whether the industry also has its own ideas about naturalness (8–9). The naturalness issue is reformulated from also, possibly, being an industry problem into a consumer concern only: it is consumers who want natural foods (14–15). In addition, the preparedness of the industry to listen to consumers is underlined. While it may not be the most logical choice to produce natural or e-number free food (more difficult, more expensive, etc. 23–25), we produce what they want. In so doing – turning naturalness into a private consumer concern which is attended to by experts (though somewhat reluctantly) – the need to further explore that concern is taken away. There is no reason for consumers to complain, so why investigate their concerns in a more than superficial manner?

Potential reasons to explore what "naturalness" refers to are further undermined by adding that consumers want natural food "without probably many consumers uhh understanding what that then means and what it entails" (16–18). This formulation defines the food expert as having superior access to what "naturalness" is, by suggesting a yardstick along which (other) definitions can be measured. By merely *implying* epistemic superiority, the actual definition of naturalness is claimed to be in the hands of experts such that there is no need to have it disclosed. Black-boxing the expert definition of naturalness prevents having it available for discussion, and opening it up – and other definitions for that matter – for debate.

Both discursive actions, i.e., framing naturalness as a private consumer-citizen concern that is already met by food experts as best it may, and claiming a superior definition of naturalness without having it explicated, work to establish naturalness as a concern that need not be dealt with in the public sphere. It is presented as already dealt with, without undermining scientific superiority or creating any pressure to ask explorative questions (as in: "What do you mean by natural food?").

While there is only space to discuss two cases here, the fragments shown here seem to represent a broader pattern in which potential soft impacts of future technologies either emerge as difficult-to-pin-down and mediated concerns (as with the celiac pill), or come up in the form of black-boxed, classic arguments (as with naturalness). While in the first case, the soft impact or identity concern is only implicitly available (as an interactional goal of participants' utterances rather than in the literal content of what they say), in the second case it is explicitly there but constructed as private and not in need of further exploration (this is again achieved indirectly, as an interactional consequence of the expert's arguments). Both ways of dealing with soft impacts make them susceptible to denial in the public domain, either because they are not visible, or because they are treated as private, known, and already dealt with (though nonvalid). The question is: how come?

## **Three Dimensions of the Hard/Soft Distinction: An Explanatory Model**

---

In our analysis we focused on two concrete cases, in which impacts regarding health, identity, lifestyle, taste, and naturalness were at stake. In the first case we showed how a patient raised her concerns about the pill's impacts on identity and lifestyle only in an indirect, roundabout way. In the previous section we saw how technology actors manoeuvered to allocate accountability for impacts. Some of these got accepted as public concerns that deserve the attention of technology actors and policy makers, while others got framed as private concerns and delegated to the citizen-consumer. These two cases seem to exemplify a wider pattern: some topics get taken up by technology actors, such as health, safety, and environment, while others, such as identity, lifestyle, and naturalness, are hardly taken serious. In this section we offer an explanatory model: in liberal, secular societies in which science and legal conceptions of accountability play pivotal roles, some of technology's impacts get qualified as "hard," others get dismissed as "soft." This crucial distinction is made along three dimensions.

### **Valuation**

---

The first dimension regards the valuation implicit in hopes and fears regarding the impacts of emerging technologies. If we look back at the examples given in the previous sections, the values underlying the concerns would be something like "having a sense of achievement," as exemplified in the diet-centered lifestyle of the celiac patient, or the "naturalness" of food. But the defenders of these values have, as the analysis shows, a hard time making themselves heard. It shows in the way Sammy talks: She *blurts* out that she doesn't need the pill, rather than "rationally" assessing the pros and cons of that particular medical technology, and without explaining how the prospect of the pill somehow affronts her. In the second example, it is clear that some consumers worry about whether modern food technology somehow results in "unnatural" food. It is equally clear, however, that this concern is not really taken seriously by the technologist. He bows for the demand, but only like an adult sometimes bows to the demand of an obstinate child: It may not be wise, but it is easier as it avoids a hassle.

Lifestyle concerns like "sense of achievement" or "naturalness" somehow seem to be taken less seriously. Nominally, there is no reason why these values would not be included in risk

assessments. Risk is simply defined as the probability that something undesirable will happen, so that could refer to any value. However, in actual practice, the values usually implied in Risk Assessment are only two: Safety and Health. True, in recent years, the Environment (Sustainability) was an important addition to the values implied in risk, and yes, more recently Privacy seems to be gaining prominence. Finally, Technology Assessment is usually broader than Risk Assessment and also takes into account values like Economic Growth and Employment. But that is about it, value-wise.

This is strange, as people have worried about a much broader palette of values in relation to technology: about the erosion of tradition, the tendency toward uniformity and conformism, about alienation, dehumanization, depersonalization, spiritual shallowness, enslavement by the machine, devaluation of life's fundamental values, artificial needs, about Faustian hubris, playing God, Frankenstein, about threats to democracy and justice, privacy, and so forth. Or they have hoped for much more important benefits: true self-development, post-humanism, true religion, world peace, cosmopolitan understanding, and so forth.

How then to explain this narrow focus of Technology and Risk Assessment? The answer lies in the dominance of liberalism in our societies. The key value informing liberalism is individual freedom, nowadays most often operationalized as "freedom of choice." The restriction of that freedom by the state is *a priori* under suspicion and always has to be justified. To this day, the simplest, most powerful, and most wide-spread justification of state intervention circulating in Western societies is J.-S. Mill's no-harm principle: "That the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others" (John Stuart Mill 1859, pp. 21–22). So, in the case that there is no clear, noncontroversial instance of harm done, liberals lose interest. Those issues are left for everyone to decide upon individually, that is, they get relocated from the public to the private domain, where they are treated as matters of subjective preference. In John Rawls' influential terminology: public reason deals with the "Right," not with comprehensive conceptions of the "Good" (Rawls 1993, pp. 173–211).

When a nuclear reactor explodes, that is harmful. No one hesitates to affirm such a statement. Hard impacts are considered to be hard because they refer to such indubitable instances of harm: a technology is good when it helps avoiding it – e.g., by providing a cure against cancer or by helping to feed the hungry – and bad when it causes such harm. In the latter case, the state should move in. Safety, Health, Sustainability, Privacy, Profit, and Employment: When technology touches upon these values, relevant actors (technologists, policy makers, citizens) agree that these impacts qualify as harm, and should therefore be recognized as matters of public concern.

But unfortunately things are not always so clear-cut. When the television pollutes our minds by producing large quantities of inane chatter, is that harmful or innocent fun? And does Internet turn our friendship into a travesty (Turkle 2010) or do our ideas about friendship simply evolve with the new technological reality? Or, to return to the examples discussed in the previous sections: is it harmful when new medication threatens to rob a particular lifestyle of its value, and the ones living it of some of their sources for self-esteem? Or when technological interventions alienate us from our food, is that bad? Some would answer yes, but many would not.

If a technology is detrimental to one's safety or (preferably physical) health, few are going to argue. But it is much more difficult to establish a broad consensus on moral, cultural, or political "harms." In a liberal, pluralist society that prides itself on its tolerance of diverging conceptions of the good life, technologies cannot be forbidden on such a shaky basis. And

because it cannot be forbidden, why talk about it at all? In liberal societies about the only harm that is considered a legitimate topic for public discussion, is physical (or maybe medically certified psychological) harm, because only on this topic citizens can reach consensus. In other words, there is an – admittedly gliding – scale between impacts that are conceived to be “hard” because they involve clear instances of harm, and impacts that are conceived to be “soft” because they do not. Technology and policy actors take the first type seriously, but rarely the latter type (Swierstra 2002).

## Quantifiability

---

Now let us turn to the second dimension of the distinction between hard and soft impacts: quantifiability. Quite apart from the kind of harm we are dealing with when assessing the impacts of a technology, we also want to know how big the chance is that a technology will cause such harm in the future, and how big the harm then will turn out to be. So, how big would be the risk that the gluten pill would indeed affect the diet-centered lifestyle of Sammy, and how harmful would that be exactly? Or: how big is the probability that modern food technologies diminish the “naturalness” of our food, and if so, how bad would this be exactly?

Both technology actors and policy makers tend to prefer answers to these questions in the form of numbers. For them, numbers equal objectivity. The more readily impacts lend themselves to quantification, as the better they fit into the discourses prevailing among technology developers and policy makers and the more readily they are accepted as “rational” and “serious,” in other words, the “harder” they are perceived to be (cf. Slovic 2000; Jaeger et al. 2001; Roesser 2010). And indeed, some risks do lend themselves to this language of numbers. An example of high quantifiability is the risk of a nuclear disaster, both in terms of probability and in terms of body count. In general, we can say that impacts on Health and Safety, and on Profit and Employment, can be quantified well using numbers. Environmental risks, on the other hand, already lend themselves less readily to quantification. Their probability may still be calculated but it often proves difficult to attach numbers to the harm/impact itself. Of course, one can estimate how many fish will die, but how to translate this quantity into a magnitude of harm – to us? This is why harms to the environment often get translated into economic terms. Risks to our Privacy are also hard to quantify.

But many of technology’s impacts lend themselves even less to quantification. Take for example the risk that a new medication will change my diet-centered life style and undermine my sense of achievement? By what means to assess the probability that that will happen, and how even to begin quantifying such an impact? Or the risk that food technology will alienate us even further from Nature. How to calculate the probability that that will happen? And how to even start measuring different degrees of alienation?

Quantifiable risks count as hard, nonquantifiable risks get dismissed as too soft to merit rational discussion. Why do technology actors and policy makers seem to prefer the language of numbers? The answer to that question is not self-evident. More than a century ago, Wilhelm Dilthey argued that there are two ways of investigating the world: scientific explanation for the natural world and (historical) understanding for the social world of meanings. But still, up to this day scientists and policy makers consider the contributions of history, anthropology, and other qualitative sciences as too soft to take seriously. Similar to the way “harm” is considered

as an objective criterion in liberalism, allowing for a rational discourse capable of generating consensus, in science and policy making “quantifiability” is perceived as a sign of objectivity and rationality.<sup>2</sup> Only on this basis a rational consensus is deemed to be possible.

## Causality

---

However, to be really accepted as “hard” by technology actors and policy makers, an impact has to meet a third and final condition. To be relevant to these actors, they have to somehow feel responsible, or more passively, afraid to be held accountable, for the impact in question. A major precondition for responsibility/accountability is that there exists a clear causal link between technology and impact.<sup>3</sup> When such a link can be established, this considerably adds to the hardness of an impact. And who would try to deny the causal link between a nuclear explosion and the dead bodies around?

But the causal link between technology and impact is not always easy to establish. Philosophy of technology, Actor Network Theory, and (post-)hermeneutics have argued convincingly that the conception of technology as a passive, neutral instrument is naïve. Philosophers of technology point out that technology is far from passive and neutral, because it mediates our (theoretical and practical) relations with the world in specific ways. Technology can change the way we interpret the world (Idhe 1993), and how we act in it (Akrich 1992; Latour 1992; Verbeek 2005; Swierstra and Waelbers 2010; Waelbers 2011). Studies show over and over again how technological artifacts, for instance, can “invite” or “facilitate” certain behavior in the user. These instances of technologically mediated behavior are, however, difficult to assess in terms of accountability. The causal link between technology and impact is not straightforward, but bent, diffused. A philosopher may consider accountability for undesirable impacts distributed over technologists, users, artifacts, and policy makers; in legal practice it is still usually the user who ends being blamed.

As American bumper stickers never tire to explain: Guns don’t kill people; People kill people. Of course, even the gun lobby is willing to admit that in some cases guns do kill people, that is, when they malfunction and explode in the face of the shooter. But in all other cases, according to the weapon-lobby, it is solely the user who is to be held accountable, not the innocent (neutral, passive) *instrument* or its designer/manufacturer/seller. Similarly, if a new anti-gluten pill threatens someone’s identity, this can never be attributed to this pill. Some people will be able to withstand the pressures of this new medical technology, and stick to their old identity, so ultimately it is a matter of free individual choice.

Now, even if we think this reasoning is a little too comfortable, we have to admit that the attribution of responsibility is difficult when it is clear that effects are co-produced by a plurality of actors. We as yet have very limited means to conceptualize and organize collective responsibility. Our dominant moral models ultimately refer back to individuals making conscious choices. In cases where humans and nonhumans share responsibility, it is easier to conclude that no one is responsible. As a result, impacts of technology that cannot be clearly and unequivocally linked to technology actors are treated as “soft” and removed from the public agenda. Do not blame the makers of the gluten pill for undermining your sense of achievement. Do not blame the food technologists for making you eat unnatural food. It is no one’s fault, really, and therefore not a matter of public concern.

## Concluding Remarks

---

We started our chapter by pointing out that the participatory agenda is managed in such a way as to deal with certain topics and not with others. In the previous section, we argued that some concerns were allowed on that agenda because they are perceived to be “hard” enough to allow for rational debate. Hard issues are essentially considered to be hard, because they promise to be the object of a rational, uncoerced, consensus: because the type of harm is noncontroversial, because no one can argue with numbers, and/or because technology or policy actors cannot deny accountability because of the clear causal link between technology and impact. Everything that does not score on (one or more of) these three dimensions, runs the risk of being dismissed as too soft – subjective, unproven, and/or messy with regard to whom is to be held accountable.

The celiac pill example showed how “soft” concerns cannot be recognized so much in what people literally say as in what becomes visible in the interactional concerns that they display, e.g., treating the anti-gluten pill – couched by an expert as a panacea to their problems – as a threat to their identity and a devaluation of their current lifestyle. This appearance clarifies why soft concerns may surface in a roundabout way rather than become apparent straightforwardly, and thus may be difficult to identify.

In other cases, soft concerns seem directly available for discussion, as in the food expert discussion on naturalness. However, this soft concern was subsequently constructed as a private consumer issue that does not require any further scrutiny or exploration, as it is already known and can be met (if somewhat reluctantly). While in this case the soft concern seems easier to recognize, the interactional result is the same: it is constructed as not deserving any further attention in the public arena.

The three dimensions of soft impacts as laid out in the previous section (difficult to value; quantify, and explain causally) make their indirect emergence or lack of exploration plausible, for this type of soft concern can expect an unwelcome reception. The patients’ talk shows an orientation to such challenge and marginalization by phrasing the rejection of the anti-gluten pill in extreme terms and not spell out the nature of the affront. But the dismissal of these concerns also requires a detour. By couching the dismissal of natural food in the obligatory language of mutual respect and of the sovereignty of the citizen-consumer’s wishes, the expert’s talk shows anticipation of the “hardness” of democratic norms and rules that demand that everyone’s concern counts.

We want to argue that this dismissal of soft impacts by technology actors and policy makers is shortsighted. It is a cause for concern when citizens fail to acquire a fair hearing for their concerns, even if the values concerned are contested, even if the chance that the harm occurs cannot be quantified, and even if there is no one who can be held accountable in a clear and unequivocal manner.

It is worth pointing out that hard impacts are not as hard as they are taken to be. There is always room for conflicts about what constitutes harm, how to quantify it, and who is to be held accountable. But more importantly, taking soft impacts seriously is not only paramount for democratic reasons – if large strata of society hold these concerns, that is in itself enough reason to discuss them carefully – it is also crucially important for substantive ones.

First of all, taking a broader range of values seriously opens a door to a more positive heuristics with regard to emerging technologies, away from the present binary discourse about the question whether a technology should be forbidden or not. Currently the main thrust in Risk or Technology Assessment is negative: How to avoid or minimize harm? If no clearly

harmful impacts are to be expected, policy makers and technology actors lose interest and the success of the technology is now left to the unreflective preferences of individual consumers. But in a technological culture like ours, the issue should rather be: how to establish a practice of public deliberation on what *good* technology is. What technology do citizens want to see developed? The aim should be goal-setting rather than harm-avoidance. Taking soft impacts seriously helps to move away from the binary risk discourse (“Should this technology be forbidden: yes/no”), to a discourse of the (common) good (Swierstra 2002).

Secondly, laying too much stress on quantifiability can be highly counterproductive. Because they only had eyes for the hard impacts of GMOs, decision makers for too long dismissed the public’s doubts as irrational, emotional, private, and religious, etc. The resulting break of mutual trust between producers and consumers has frustrated the development of biotechnology (Wynne 2001).

Thirdly, ignoring indirect impacts may thwart the technology’s intended aim, as in the case of so-called revenge-effects (Tenner 1996). Technology actors and policy makers often fail to anticipate that the user’s behavior changes because of the new technology. The “light” cigarettes that in the end only increased the net intake of tar and nicotine because people assumed these were not so unhealthy, provide a good example. Such technologically mediated behavioral change is currently dismissed as a soft impact, because responsibility cannot be unequivocally located with the technologist. But when such indirect impacts are foreseeable for the marketing department of the tobacco company, why should technology developers and policy makers be excused from taking them into account?

Summarizing: In the case of controversial technologies, like the life sciences for instance, stakeholders point out a large array of possible impacts. However, decision makers, like; technology actors and policy makers, tend to concentrate on “hard” – quantifiable, harmful, direct – impacts. But it is essential that in a technological culture soft impacts of emerging technologies are equally taken into account. This is the only way to make the co-evolution (Rip and Kemp 1998) or co-production (Jasanoff 2004) of technology and society reflective and democratically checked. Technology does far more than simply providing the means to our existing goals. Technology redefines these goals; changes or affirms power relations; affects values, standards, and norms; informs aspirations; installs new needs and preferences; teaches what it is right to hope for.

## Further Research

Having said that responsible innovation processes require soft impacts to be taken into account, it is important to point out what we mean by the latter, and whose responsibilities we are and are not referring to. For one thing, “taking into account” soft impacts does not equal accepting these impacts as true or right, and/or following them up immediately. We do not propose that expert-technologists start to grow natural food straight away, or that policy makers acknowledge that genetic modification transgresses ethical boundaries once citizens have pointed those out to them. The validity of soft impacts, and acting according to them, should become part of the negotiation. Furthermore, while technologists and policy makers may be inclined to display little concern for soft impacts, this neither implicates that it is their responsibility alone to solve the matter, nor that it would be the most effective way to go.

As the conversation analysts Heritage and Raymond (2005, p. 2) point out: “the distribution of rights and responsibilities regarding what participants can accountably know, how they know it, whether they have rights to describe it, and in what terms, are directly implicated in organized practices of speaking.” Looking at the actual dynamics of accountability attribution and denial, particularly at the level of what expert participants claim to accountably know, i.e., claim to have access to, and for what interactional purposes, is crucial as a first step for revealing the ways in which the hard/soft distinction are made and sustained. Not only as to understand better the concerns that guide, consciously or not, the referral to soft and/or hard impacts, but also because different attributions of accountability may have different implications for how to achieve a more comprehensive public dialogue.

That naturalness is both constructed as a private consumer concern and black-boxed as requiring no further investigation, makes it a different problem than, for example, the issue of good taste (Middendorp et al. *in prep.*). Food experts tend to attribute complete responsibility to consumers for telling them what good taste is, and only claim epistemic access to the technicalities of how to achieve a certain taste. Since anything can be made – is the suggestion – there is no such thing as a lack of good taste when it comes to future foods. In both cases, soft concerns are pushed off the discussion table but the starting point for a more comprehensive dialogue would be different. Naturalness and good taste are both treated as private preferences that require no debate, but taste is constructed as a legitimate concern, whereas naturalness is dismissed as invalid. For one thing, naturalness would need to be explored, and the conclusion may well be – jointly with consumers – that “natural food” is infeasible, or precisely the reverse, because it stands for something other than expected. Up till then, experts keep the “problem” intact as much as they blame consumers for.

Our analysis shows that these concerns are often not or only indirectly available for the debate facilitator. Likewise, participants themselves tend not to have direct access to interactional concerns although, when confronted with them, they will recognize them immediately. The Discursive Action Method (Lamerichs and te Molder *forthcoming 2011*) is a reflection method that aims to turn participants into analysts of their own discourse by making these concerns visible and open for discussion. This not only counts for the expert-designer or policy makers but just as well for (potential) users of technologies. Natural food may be reshuffled into a private consumer concern with which a food expert should not be preoccupied, but “naturalness” may just as well be drawn upon by consumers to delineate their territory such that no expert is allowed in.

A close and critical reflection on how soft/hard concerns are drawn upon, and for what interactional business, may be the starting point for a new area of research, and a practice in which a more comprehensive dialogue could make a start. This research should then be complemented by a philosophical critique of the three dimensions that together make up the hard–soft distinction. Such critique will have to draw its inspiration from quite diverse traditions. The primacy of the no-harm principle in liberal political philosophies has to be investigated in the light of the new realities of a technological culture. Does the way this principle is applied allow for fruitful public deliberations about the (un)desirability of technologies? A similar investigation has to focus on the widespread belief that only numbers allow for rational consensus. Part of this investigation will be primarily philosophical in character, but important inputs are also to be expected from more empirical research in Science and Technology studies that explore how these numbers are constructed and contested. Last but not

least, the issue of technological mediation has to be explored further by both philosophers of technology, who investigate the various forms of technological mediation, and moral philosophers, who have to develop convincing conceptions of collective, or shared, responsibility. These types of research in the philosophical foundations of the hard–soft distinction will help to create the necessary discursive space for the technologists, policy makers, and citizens. Because they are the ones who have to make sure that in their mutual dealings the (implicit) distinction between hard and soft impacts no longer serves to remove relevant topics from the agenda for the public dialogue on technology.

## Notes

---

1. Transcripts employ the notational convention used in conversation analysis (Jefferson 2004). The transcription symbols used here are:

bu-	a cut-off or self-interruption
↑	sharp rise in pitch
(1.0)	numbers denote silence in tenths of seconds
wants	underlined items were hearably stressed
(( ))	transcriber's description of events

The fragment is translated from Dutch to English, remaining as close as possible to the original Dutch text.

2. A separate issue, of course, is whether all the relevant data are available. The precautionary principle is a procedural rule devised to deal with such a (temporary) lack.
3. Although it has to be admitted that in the case of positive impacts, this demand for a direct causal link is usually interpreted less strictly. As Ravetz famously put it: "Science takes credit for penicillin, while Society takes the blame for the Bomb" (Ravetz 1975, p. 46).

## References

---

- Achterhuis H (ed) (2001) American philosophy of technology: the empirical turn. Indiana University Press, Bloomington/Minneapolis
- Akrich M (1992) The description of technical objects. In: Bijker W, Law J (eds) Shaping technology, building society: studies in sociotechnical change. MIT Press, Cambridge
- Boenink M, Swierstra T, Stemerding D (2010) Anticipating the interaction between technology and morality: a scenario study of experimenting with humans in bionanotechnology. *Stud Ethics Law Technol* 4(2): article 4
- Edwards D (1997) Discourse and cognition. Sage, London
- Heritage J, Raymond G (2005) The terms of agreement: indexing epistemic authority and subordination in talk-in-interaction. *Soc Psychol Q* 68(1):15–38
- Hobson-West P (2007) 'Trusting blindly can be the biggest risk of all': organized resistance to childhood vaccination in the UK. *Sociol Health Illn* 29(2):198–215
- Idhe D (1993) Postphenomenology. Northwestern University Press, Evanston
- Jaeger CJ, Renn O, Rosa EA, Webler T (2001) Risk, uncertainty, and rational action. Earthscan, London
- Jasanoff S (2003) Technologies of humility: citizen participation in governing science. *Minerva* 41:223–244
- Jasanoff S (ed) (2004) States of knowledge: the co-production of science and social order. Routledge, New York
- Jefferson G (2004) Glossary of transcript symbols with an introduction. In: Lerner GH (ed) Conversation analysis: studies from the first generation. John Benjamins, Amsterdam/Philadelphia, pp 13–31

- Lamerichs J, te Molder H (2011, frth) Reflecting on your own talk: the discursive action method at work. In: Antaki C (ed) *Applied conversation analysis. Intervention and change in institutional talk*. Pallgrave Macmillan, Basingstoke
- Latour B (1992) Where are the missing masses? In: Bijker W, Law J (eds) *The sociology of the new mundane artefacts. Shaping technology, building society*. MIT Press, Cambridge
- Marris C (2001) Public views on GMOs: deconstructing the myths. *EMBO Rep* 21(7):545–548
- Middendorp S, te Molder H, van Woerkum C (in prep.) Responsible innovation in the food sector: what impacts of food technology may enter the public debate? Wageningen University, Wageningen
- Mill JS (1859) *On liberty*. Oxford University, Oxford, pp 21–22
- Potter J (1996) *Representing reality. Discourse, rhetoric and social construction*. Sage, London
- Ravetz JR (1975) ...et augebitur scientia. In: Harré R (ed) *Problems of scientific revolution. Progress and obstacles to progress in the sciences*. Clarendon, Oxford, pp 42–57
- Rawls J (1993) *Political liberalism*. Columbia University Press, New York
- Rip A, Kemp R (1998) Technological change. In: Rayner S, Malone EL (eds) *Human choice and climate change*, vol 2. Battelle, Columbus, pp 327–399
- Roesser S (ed) (2010) *Emotions and risky technologies*. Springer, Dordrecht/London
- Slovic P (2000) *The perception of risk*. Earthscan, London
- Swierstra T (2002) Moral vocabularies and public debate: the cases of cloning and new reproductive technologies. In: Keulartz J, Korthals JM, Schermer M, Swierstra T (eds) *Pragmatist ethics for a technological culture*. Kluwer Academic, Deventer, pp 223–240
- Swierstra T, Waelbers K (2010) Designing a good life: the matrix for the technological mediation of morality. *Eng Ethics* (Online First, 30 Nov 2010)
- Swierstra T, Stemmerding D, Boenink M (2009) Exploring techno-moral change. The case of the obesity pill. In: Solllie P, Duwell M (eds) *Evaluating new technologies*. Springer, Dordrecht, pp 119–138
- te Molder H (2008) Discursive psychology. In: Donsbach W (ed) *The international encyclopedia of communication*, vol IV. Wiley-Blackwell, Oxford, UK/Malden, pp 1370–1372
- te Molder H, Potter J (eds) (2005) *Conversation and cognition*. Cambridge University Press, Cambridge
- te Molder H, Bovenhoff M, Gremmen B, van Woerkum C (submitted) Talking future technologies: how celiac disease patients neither accept nor reject a ‘simple pill’
- Tenner E (1996) *Why things bite back. Technology and the revenge of unintended consequences*. Knopf, New York
- Turkle S (2010) *Alone together. Why we expect more from technology and less from another*. Basic Books, New York
- Van der Pot JHJ (1985) *Die Bewertung des technischen Fortschritts. Eine systematische Uebersicht der Theorien*. Van Gorcum, Maastricht
- Veen M, Gremmen B, te Molder H, van Woerkum C (2010) Emergent technologies against the background of everyday life: discursive psychology as a technology assessment tool. *Public Underst Sci*. doi:10.1177/0963662510364202. Prepublished 13 Apr 2010
- Verbeek PP (2005) *What things do. Philosophical reflections on technology, agency, and design*. Pennsylvania State U.P., Pennsylvania, PA
- Waelbers K (2011) *Doing good with things—taking responsibility for the social role of technologies*. Springer, Dordrecht
- Wynne B (1996) Misunderstood misunderstandings. Social identities and public uptake of science. In: Irwin A, Wynne B (eds) *Misunderstanding science? The public reconstruction of science and technology*. Cambridge University Press, Cambridge, pp 19–46
- Wynne B (2001) Creating public alienation: expert cultures of risk and ethics on GMOs. *Sci Cult* 10(4):446–481
- Wynne B (2006) Public engagement as a means of restoring public trust in science – hitting the notes, but missing the music? *Community Genet* 9:211–220

# 43 Risk and Technology Assessment

Rinie van Est · Bart Walhout · Frans Brom  
Rathenau Institute, The Hague, The Netherlands

<b>Introduction .....</b>	<b>1069</b>
No Straightforward Connection .....	1070
Complementary Political Roles .....	1070
Contents .....	1071
 <b>Classical Risk Approach and TA .....</b>	 <b>1072</b>
Classical Risk Approach .....	1072
Risk as a Calculable Property .....	1072
Separating Science and Politics .....	1073
Classical TA as Expert-Driven Policy Analysis .....	1073
Empowering Congress vis-à-vis Government .....	1073
Policy Analysis Versus Democratization .....	1073
Comparative Conclusions .....	1074
Two Basic Shortcomings of the Classical Risk Approach .....	1075
 <b>Interpretation of Risk with RA and TA .....</b>	 <b>1075</b>
RA: From Calculable to Uncertain Risks .....	1075
Bad and Good Chances .....	1076
Public Perception .....	1076
Complexity .....	1076
Broader Public Concerns .....	1076
Risk Versus Uncertainty .....	1076
Uncertain Risks .....	1077
TA: From Social Effects to Drivers .....	1077
Bad and Good Chances .....	1077
New Types of Side Effects .....	1078
Social Drivers .....	1078
The Boundaries of TA .....	1078
Comparative Conclusions .....	1078
 <b>Risk Characteristics and RA and TA .....</b>	 <b>1079</b>
Risk Situations and Risk Management Strategies .....	1079
Characterizing Problem Situations and Type of TA .....	1080
Comparative Conclusions .....	1081

<b>New (Participatory) Approaches to TA and RA .....</b>	<b>1082</b>
New (Participatory) Approaches to TA .....	1082
Improving the Interaction Between TA and Politics .....	1082
TA as an Interface Between Politics and Society .....	1083
New (Participatory) Risk Approaches .....	1083
The IRGC Risk Governance Framework .....	1084
Risk Assessment Plus Concern Assessment .....	1084
Improving the Interface Between the Risk Assessment and Management Sphere .....	1085
Comparative Conclusions .....	1085
 <b>An Example: Risk Governance of Nanotechnology in the Netherlands .....</b>	<b>1086</b>
Setting the Agenda .....	1086
Setting Up the Risk Governance Machinery .....	1087
Dealing with “Organized Irresponsibility” .....	1087
Monitoring and Debating Risk Governance .....	1088
Parliamentary TA’s Role in Risk Governance .....	1088
 <b>Further Research .....</b>	<b>1089</b>

**Abstract:** This chapter provides an overview of the changing relationship between risk, technology assessment (TA), and risk assessment (RA). It does so by comparing the development of the practice of parliamentary TA and RA, the way risk is interpreted in these practices, and the political role these practices play in dealing with risks. The basic argument is that originally RA and TA presented politically separate practices. Over the last decade, the conceptual gap between these two practices has been bridged to a large extent. We start with describing the classical approaches to TA and RA, which developed in 1960s in the United States and were guided by the belief that scientific methods would improve decision making around the risks involved in science and technology. Classical parliamentary TA and RA present very distinct scientific and political practices, with different conceptions of risk and political roles. The classical approach to risk operated with a narrow mathematical definition of risk. Classical TA defined risk in a much broader fashion; risk referred to a broad set of (potential) negative social effects of science and technology. RA was thought to help the government in managing risk, by depoliticizing risk management. In contrast, parliamentary TA aimed to enable a political debate within Congress, and thereby strengthening the position of Congress vis-à-vis the executive branch. Throughout the years, both practice and scientific literature have revealed basic shortcomings of the classical approach to TA and risk. Driven by the concept of uncertainty, the role of RA and TA and their interpretation of risk have changed. Modern risk approaches are expected to deal with both calculable and uncertain risk. TA is encouraged to look beyond effects, to also analyze current visions and values that drive science and technology. Based on the concept of uncertainty, attempts have been made to characterize risk or problem situations in order to clarify the limitations of the classical RA and TA approaches. The claim is that in case of scientific and regulatory uncertainties, and value dissent more participatory approaches to RA and TA are required, which seek to represent public controversy. The IRGC risk governance framework can be seen as exemplary for the new risk approach. From a risk governance perspective, RA and parliamentary TA have become complementary practices. The case of risk governance on nanotechnology in the Netherlands proves this point. However, parliamentary TA's role within risk governance presents a remarkable blind spot on the current research agenda.

## Introduction

---

This chapter aims to discuss the relationship between risk and technology assessment, in particular parliamentary technology assessment. Technology assessment (TA) is driven by an awareness about potential positive and negative effects of technological change, and the hope that one can anticipate these effects. Since it deals with the interplay between technological change and (potential) social problems, TA has a clear political side to it (Van Est and Brom 2012). This counts in particular for parliamentary TA, which this chapter will focus on. Although risk and TA clearly touch each other in all kinds of ways, such an exercise – to our knowledge – has never been undertaken before. This is quite surprising since on first sight, the relationship between risk and TA seems to be a close and obvious one. Take, for example, this quote by Grunwald (2009, p. 1131): “One of the main reasons for the emergence of TA was because of the risks directly or indirectly caused by technology and its use... TA should and does contribute to the early signaling of risks and how they should be dealt with.” Discussing the above relationship, however, is a rather complex, but still worthwhile task, because it is of particular interest for current debates on risk politics and governance.

## No Straightforward Connection

---

Discussing the relationship between TA and risk is not a straightforward task. Take for example, the fact TA is not built around a precisely defined concept of risk. At least the classical risk approach provides a clear technical definition of risk, even a formula to calculate risk. Namely risk is generally defined as the product of the magnitude of the possible adverse consequence(s) and the probability of occurrence of each consequence. In contrast, “risk” within the current TA practice tends to refer to a “colloquial meaning of the term risk” (Maasen and Merz 2006, p. 25), in the sense that risk in TA equates with (negative) social impacts of science and technology in general. There are at least three other reasons why the connection between risk and TA is not obvious.

First, no unambiguous and selective definition of TA exists (Grunwald 2009). This is because TA is neither a separate field of scientific research nor a well-defined, clear-cut practice. Disciplines ranging from policy and political sciences to ethics, science, and technology studies; communication sciences; and social and cultural studies have all influenced the way TA is understood, performed, and institutionalized. TA is being employed in a wide variety of institutional settings, covering many functions, goals, methods, and target groups. TA can be found in industry to help product development, close to politics to provide information, and stimulate the parliamentary and public debate. TA is also employed as an instrument to guide scientific research from a societal perspective. For example, studies on ethical, legal, and social aspects (ELSA) of science and technology have been increasingly integrated in large research programs in order to integrate societal considerations in research choices.

Second, risk assessment (RA) is about assessing (technical) risks. RA provides scientific input into the risk management process, that is, the political decision-making process about how to deal with risk. Parliamentary TA is not primarily about assessing (technical) risks. Its main political role is to help the democratic system to deal with (potential) public controversies on science and technology. These public controversies might be driven by concerns on safety, but also wider public concerns. Public controversies come into play because various groups may have different interests or different views on what the problem is, and what kind of solutions should be strived for. Democratic politics is a way to deal with public controversies that includes free and open deliberation and also exercising power. Political practice is about connecting these two poles (cf. Jaspers 1965), and TA is supposed to play a positive and constructive role in achieving that.

Finally, the practice of parliamentary TA and the practice of assessing and politically managing risk are not static ones, but constantly developing. Both the classical approach to TA and risk were first developed and institutionalized in the United States during the 1960s and the 1970s. Since then, both practices have spread around the world. Throughout the years, both practice and academic literature as well as encountering new political cultures have changed the practice of parliamentary TA. Also the classical risk approach has received a lot of criticism. In particular, the political value of the technical definition of risk has been hotly debated over the years.

## Complementary Political Roles

---

The fact that analyzing the relationship between risk and TA is dynamic and complex does not imply that it is impossible or not useful to do so. We believe that it can be fruitful to clarify in

a systematic way how parliamentary TA interprets risk and how this practice touches the practice of risk assessment. Both TA and RA provide scientific input into the political decision-making process on how society handles risks. One might say that TA relates to the parliamentary debate in the same way as RA relates to the political process of governing risks. Lynn's (1981) game metaphor of the political decision-making process gives us a second impression of how the practice of TA and RA relate to each other. Lawrence E. Lynn distinguishes three games within the political decision making, corresponding to three levels within the political system. The high game involves deciding on whether there is a role for the government. It focuses on deciding the right thing to do and identifying the social values that are explicitly at stake. The middle game is more concretely about what the role of the government is going to be, about the effectiveness and efficiency of policy interventions. The low game is about the formulation of the precise design of the related policy instruments. Roughly speaking, parliamentary debate and TA are situated in the high and middle political game, while risk assessment and management are situated more in the middle and low game. The roles and practices of TA and RA thus seem to be complementary, and partially overlapping. From this line of argument follows the central aim of this paper: to analyze to what extent the political practice of (parliamentary) TA and risk assessment may inspire and complement each other in a constructive manner. For this, we will describe and compare the political role TA and RA play within the political decision-making process on risks, in connection to the way they interpret the notion of risk.

## Contents

---

The first section compares the classical approach to TA and RA. Both practices arrived in the 1960s and were politically legitimized in similar ways. Both approaches were based on the assumption that science-based expertise would rationalize the political decision-making process on technology-related risks and controversies. Nevertheless, their political roles and their conception of risk differed strongly. While RA was meant to rationalize the political debate on technological risks, parliamentary TA's role was to strengthen the position of the parliament vis-à-vis the government. Besides, while RA had a narrow probabilistic definition of risk, TA used a common sense notion of risk, referring to a broad set of (potential) social effects of science and technology. The classical forms of RA and TA became severely challenged by two fundamental and interrelated forms of critique: the issue of problem framing and representation. The next three sections deal with these two aspects of problem framing.

The second section analyzes various shifts in the interpretation of risk within TA and RA. TA has broadened its scope from focusing on (potential) effects of technology toward the current visions and values that shape science and technology. RA has broadened its perspective from risks that are calculable based on past experience toward (future) uncertain risks. The third section looks at the relationship between characterizing the risk situation and its implications for the appropriate roles and approaches to RA and TA. With respect to characterizing the risk situation the notion of uncertainty has slowly entered the minds of political decision makers during the 1990s. This awareness revealed the boundaries of the classical risk approach, by limiting it to so-called simple risk problems. Moreover, it underpinned the legitimacy of participatory approaches in TA and RA.

The fourth section deals with the issue of representation, that is, the question of who is allowed to define the risk problem at stake. The basic complaint is that expert-based classical

approaches do not sufficiently represent public concerns on risks. To address this shortcoming, new (participatory) approaches to TA and RA have been proposed which aim to involve a broader set of social actors within the political decision-making process on technology-related risks. In particular, this section investigates the relationship between the risk governance model (see also ➤ [Chap. 44, Risk Governance](#) by Hermans, Fox, and van Asselt in the present book) and TA. Our analysis shows that the risk governance model includes various typical TA elements and visions. Exactly this makes a comparison of the practice of politically dealing with risk and parliamentary TA of current interest.

The fifth section describes the case of risk governance on nanotechnology in the Netherlands. This case shows how parliamentary TA has a fruitful role to play in the governance of risk. The final section will make some concluding remarks and describes some themes that require further research.

## Classical Risk Approach and TA

---

Risk assessment and technology assessment were first developed in the United States (Bimber and Guston [1997](#)). During the 1960s, the rise of decision theory, operations research, and system theory had raised the hope that scientific methods could improve decision making around science and technology (Bereano [1997](#)). This belief rendered political legitimacy and support to the idea of risk assessment and technology assessment in the United States.

For example, to calculate the probability of accidents in the field of nuclear power and aerospace quantitative risk assessment tools were developed. This led to the so-called classical risk approach. In the same period, the scientific community in close interplay with the Congress developed a classical TA approach, which defined TA as “a policy study designed to better understand the consequences across society of the extension of the existing technology or the introduction of a new technology with emphasis on the effects that would normally be unplanned and unanticipated” (Coates [2001](#), p. 303). Eventually, this led to setting up a TA institute in the U.S. Congress in the early 1970s. This section gives a short description of both the classical risk approach and classical TA, and compares the way they interpret risk, and their presumed political roles within the decision-making process.

### Classical Risk Approach

---

#### Risk as a Calculable Property

In the scientific literature many definitions of risk can be found (for an overview cf. Renn [2005](#), pp. 141–142). Most definitions, however, normally consist of two elements. The first element refers to damage, undesired impacts, or adverse effects. The second element speaks of chance, likelihood, or probability of such harmful consequences. The most prevailing definition combines the chance that a given hazardous effect will occur and the impact this will have. Risk became interpreted as the product of the magnitude of the potential loss or damage and the probability that that loss will occur. This is reflected in the often used (risk) formula: “risk = probability × effect.” The corner stone of the classical approach, thus, is to quantify certain risks.

## Separating Science and Politics

This formula forms the basis for the classical risk approach which became dominant during the 1960s in the way Western societies politically deal with risks. In this approach, basically, two stages can be distinguished. In the risk assessment phase, risk experts try to quantify relevant risks (Renn 2005, p. 27). Risk experts first have to identify and, if possible, make an estimation of the hazard. Next, they have to assess the exposure and/or vulnerability to the danger. Finally, they have to make an estimation of the risk based on the former two steps by using the risk formula. In the risk management stage, risk managers take measures to deal with or control various unacceptable risks. The classical risk approach is founded on a clear (institutional) distinction between the risk assessment and management stages, or in other words, between the “science” of estimating risks and the “politics” of taking risk measures (National Research Council 1983). It was feared that otherwise political pressure could harm scientific independence, which could lead to over- or underestimating certain dangers.

## Classical TA as Expert-Driven Policy Analysis

---

### Empowering Congress vis-à-vis Government

Over the 1960s, public awareness grew of potential health and environmental risks related to new technologies. It was felt that representative institutions were failing to deal with the negative side effects of technological change. The need to estimate the environmental impact of a proposed (technological) project became regulated by the National Environmental Policy Act (1970). TA was “in some ways an enlarged version of the awareness that lead to the extensive development of environmental impact statements (EIS)” (Coates 2001, p. 303). In 1972, the Technology Assessment Act acknowledged the need to anticipate on a broader spectrum of social effects of technological change, besides health and environmental impacts (U.S. Congress 1972). Its passage in the House led to the creation of the Office of Technology Assessment (OTA) in the U.S. Congress. The foundation of OTA was driven by the U.S. Congress’ desire to assess the political, economic, and social aspects of technological change independently from the executive. Already, in the early 1960, the Congress of the United States was confronted with an increasing science and technology budget, related to ambitious projects, like putting a man on the moon. Congress became concerned about its lack of ability to evaluate matters of scientific and technical complexity. Lacking critical, independent information, Members of Congress feared they were becoming “the rubber stamps of the administrative branch of government” (Democrat George Miller 1961 quoted in Kunkle 1995).

### Policy Analysis Versus Democratization

Several competing political agendas concerning the function of TA were in play. While some saw OTA primarily as an instrument to strengthen the Congress’s scientific oversight of federal science and technology initiatives, others wanted TA to be an instrument for democratization (Bereano 1997). The 1960s had seen the revival of Jeffersonianism, with its trust in community self-reliance and grassroots democracy. With respect to the practice of TA, Jeffersonian ethics implied a call for more public participation in the political decision-making process

concerning new technologies. In 1973 and 1974, a coalition of public interest groups and individuals tried to assure that OTA would develop participatory processes. This, however, did not become common practice within OTA. In the early 1960s, members of Congress realized that the executive branch had close relationship with the National Academy of Sciences, while Congress was lacking such contacts (Kunkle 1995). During the 1960s, closer bonds between the National Academy of Sciences and Congress were created. In that process, the scientific community was successful in turning OTA primarily into a scientific instrument. Accordingly, OTA defined its practice of TA primarily as a form of critical-rational policy analysis. The aim of this expert-driven policy analysis was seen as “speaking truth to power.” Nevertheless, Congress continued to call for inclusion of wider social values and perspectives in TA. This led to the development of a system of controlled consultation with some experts and interest groups.

## Comparative Conclusions

Both the classical risk approach and classical (parliamentary) TA are guided by the belief that scientific methods can improve political decision making around science and technology. Still, they clearly present very distinct scientific and democratic practices, with different conceptions of risk and different political roles (see ➤ *Table 43.1*).

Within the classical risk approach, risk became mathematically defined through the risk formula. Within TA risk was interpreted in a much broader and more social science oriented fashion. Although RA was part of the TA methods toolbox, “risk” within the TA community referred to everything that forms a danger. This definition reflected the demand from Congress, which was interested in a broad set of (potential) social effects of science and technology.

Related to this, the political roles of RA and TA were quite distinct. Risk assessment was thought to help the government in managing risk. The hope was that the scientific information delivered by RA could depoliticize risk management and make it into a bureaucratic process. The objective of classical TA was to provide neutral expert advice, which would strengthen the Congress’s position vis-à-vis the executive branch. In other words, its role was to enable and

■ **Table 43.1**  
Comparing the classical risk approach and classical parliamentary TA

	Classical risk approach	Classical parliamentary TA
Science	Risk assessment	Technology assessment
Interpretation of risk	$\text{Risk} = \text{Probability} \times \text{Effect}$	Broad set of (potential) social effects of science and technology
Political role		
Interface science – policy making	Informing (governmental) risk managers about technical risks	Informing MPs on social effects of S&T
Interface government – parliament	Rationalizing the political decision-making process on risk management	Empowering the Congress vis-à-vis the government
Interface society – policy making	–	Rationalizing (preventing) public controversy (by anticipating on social effects)

strengthen the political debate about risk. Also, the fact that TA was intended to assess a broad set of social effects gave Congress the means to discuss risks related to science and technology from a broad public perspective.

## Two Basic Shortcomings of the Classical Risk Approach

---

Over the last decades, both practice and scientific literature have revealed many shortcomings of the classical approach to TA and risk. Basically, two closely related categories of critique can be discerned. The first type of concerns relates to problem framing. Section [● Interpretation of Risk with RA and TA](#) deals with the issue of how to interpret risk, and what this implies for the object of study and role for RA and TA. Section [● Risk Characteristics and RA and TA](#) looks at the discussion on characterizing risk or TA problem situations. In particular, this debate plays a role in clarifying the limits of the classical approaches and legitimizing new (participatory) approaches to RA and TA.

The second category of critique is about the question of “who defines,” causes or is affected by or is responsible for dealing with the problem. It was described above that from the onset parliamentary TA was confronted with a competing participatory perspective on parliamentary TA. The classical risk approach received similar points of critique. The complexity of assessing risks was thought to give room to experts to bring in their own values in the decision-making process in a way that could not easily be detected by the public. Such an opaque highly expert-driven assessment was found not to be in tune with modern democratic principles. Since risks are often problematic complex constructs, it is a major democratic challenge to translate these constructs into “debatable risks” (WRR 2008, p. 113). Section [● New \(Participatory\) Approaches to TA and RA](#) describes and compares the rationales for new (participatory) approaches to TA and RA.

## Interpretation of Risk with RA and TA

---

Throughout the years, both RA and TA practices have been criticized for having a too narrow focus, and therefore for being scientifically misdirected or politically biased. For example, both TA and RA were accused of focusing too much on bad chances and too little on good chances. Moreover, classical risk assessment was criticized for employing a too narrow definition of risk, i.e., risk as a calculable entity. It was feared that such a narrow conceptualization would preclude a debate about broader public concerns. Classical TA was less vulnerable to such a critique because it employed a common sense definition of risk, encompassing all kinds of social effects of science and technology. Still, classical TA was criticized because of its sole focus on “social effects,” and not on “social drivers” of technological change.

## RA: From Calculable to Uncertain Risks

---

The classical risk approach assumes that it is possible to define and assess risks. The assumption that risks can be objectified and calculated has met with a lot of criticism. Notions like complexity and uncertainty to characterize the risk situation have played a central role in clarifying the limits of the classical risk approach.

## Bad and Good Chances

In *Searching for safety*, Wildavsky (1988) argued that “playing it safe” may also present a danger. He argued that besides bad chances, there are also good chances. Risk assessment then should also be able to characterize the good chances.

## Public Perception

Already in the 1970s, the assumption that risks can be “objectified” and calculated met with a lot of criticism. The classical risk approach was attacked for not being able to take account of human perceptions (cf. Slovic 1978/2000). For example, the risk formula might attribute equal weight to a major disaster having a low probability and a small accident having a high probability. Most people, however, are more fearful of a major disaster with a low probability. This critique challenges risk assessors and risk managers to come up with an approach that takes into account the way people talk about and perceive risks.

## Complexity

Moreover, it was argued that in many circumstances, it is not self-evident to define what the hazards, their probabilities, and the consequences precisely are (cf. Fischoff et al. 1981). In particular, the interactions of humans and/or technological subsystems are much more complex than an a priori risk assignment can capture. Perrow (1984) even talked about “normal accidents,” arguing that when a technology has become sufficiently complex and tightly coupled, accidents are inevitable and therefore in a sense “normal.” As a consequence, the role played by organizational failures are very hard to take into account. This also counts for the human element in the decision-making process around risk.

## Broader Public Concerns

It was also feared that a narrow focus on the assessment of risks, but also “risk perception,” would prevent a debate about broader public concerns (Felt 2007). For example, with respect to GM food, the classical risk approach focuses on assessing the related safety risks. In the public debate, however, freedom of consumer choice plays a central role. Another major source of public concern is the inadequacy of the classical approach to deal with risk. The classical risk approach does not provide space for a proper discussion about the institutional incapacity to deal with (often unpredictable) ethical and social impacts of science and technology. As a consequence, the shortcoming of the classical risk approach exactly provokes such a debate.

## Risk Versus Uncertainty

The concept of uncertainty has played an important role in clarifying the limits of the classical risk approach and promoting new approaches and considering wider public concerns. It has been argued for long that not all decision-making situations can be characterized as situations of risk. The economist Knight (1921) made a distinction between situations characterized by

risks and uncertainty. In some circumstances, entrepreneurs may be able to calculate certain risks based on experience. In other cases, decision making is more speculative. In those cases, Knight talked about making decisions under uncertainty. Harremoës et al. (2001, p. 192) have emphasized a third type of problem situation: ignorance. According to the authors, risk is about “known” impacts and probabilities, and a situation of uncertainty is characterized by “known” impacts, but “unknown” probabilities. One may speak of a situation of ignorance when also the impacts are “unknown.”

## **Uncertain Risks**

Various authors have argued that the above dichotomies – risk versus uncertainty, calculable versus non-calculable, and knowing versus non-knowing – are flawed (cf. WRR 2008; Van Asselt et al. 2009). For several reasons, they prefer to talk of “uncertain risks” (Everson and Vos 2009). First of all, such a term better connects to the way society speaks about and deals with risks. In the public debate various types of uncertainty – e.g., scientific uncertainty and regulatory uncertainty (Hood et al. 2001) – are attributed to the notion of risk. Here, risk broadly refers to “bad chances,” like damage, loss, calamities, and disasters. Moreover, despite all the uncertainties involved in waste disposal, genetically modified food, or climate change, these developments are subjected to RA and management. Second, risk situations characterized by uncertainty have become increasingly common. This relates to the fact that current risk assessment is mostly future-oriented. The basis for risk assessment, therefore, has shifted from *probability*, based on experience in the past, to *possibility*, based on expectations about the future. Finally, these authors prefer to speak of “uncertain risks” to imply that risk situations of uncertainty do not make science and expertise irrelevant. Although most risk situations are characterized by uncertainties, they are certainly not characterized by the absence of knowledge. Within the classical risk approach the role of science was “to speak truth to power.” Uncertain risks imply a different role. As Van Asselt et al. (2009, p. 363) hold: “In the context of uncertain risks, risk assessment has, or should have, a different meaning: delineating uncertainty information seems an important challenge.”

## **TA: From Social Effects to Drivers**

Classical TA was criticized for focusing too much on the potential negative effects of science and technology. Moreover, ethical and policy analysis have stimulated TA to focus on novel types of side effects and the social visions and values that are shaping the technology.

## **Bad and Good Chances**

From the start, some American politicians were hostile to TA, based on the perception that it would automatically imply regulation of technology. In the early 1970s in America, critics of government intervention in the innovation process, therefore, derided the concept as “technology arrestment” or “technology harassment” (Kunkle 1995). Developing a “constructive” view of TA substantially contributed to the growing acceptance of TA in the German Bundestag

(Paschen 2000, pp. 102–103). The institutionalization of TA in Germany was driven by a need to rationalize the debate on science and technology. One way to rationalize the debate was to also focus on the positive aspects of technological change and explicitly explore its social, economic, and ecological possibilities. This also politically legitimated German parliamentary TA.

## New Types of Side Effects

Over the last two decades, ethical analysis has become more and more integrated in the practice of TA. The ethics view opened up the debate on what risks or aspects should be taken into consideration. Taking deontological perspectives and discussions regarding the good life and the good society into account, broadens the TA agenda. Ferrari and Nordmann (2009, p. 56), for instance, propose to expand the notion of “risk.” In addition to economic, scientific, or technological benefit/risk analysis, they promote a “philosophical hope/risk analysis.” Besides, ethical analysis challenges TA to assist society in reflecting upon the possibility that technologies, like augmented reality and brain implants, may have an effect on our morals and ethical vocabulary (Swierstra et al. 2009).

## Social Drivers

Moreover, the ethical perspective forces TA to take into consideration the (variety of) deep core values of people that are currently at stake in the debate on technology. A similar plea comes from policy analysis. The argumentative turn in policy analysis presents a shift away from a rational decision making model of politics (politics of interest model) toward a more constructivist approach to policy making (Fischer and Forester 1993). Within this so-called politics of meaning model policy makers are guided by their policy belief systems and political decision making is examined in terms of interacting belief systems. The argumentative turn in policy analysis has stimulated the development of new TA methods, notably interactive TA (Grin et al. 1997) and vision assessment (Grin and Grunwald 2000). Starting from an analysis of the background theories of the various involved actors, the main challenge is to develop joint constructions and, ultimately, lines of action.

## The Boundaries of TA

Finally, classical TA is founded on the belief that one can anticipate on the various effects technology has on society. Ethical analysis questions this central assumption behind TA that one can influence science and technology. It, thus, forces the practice of TA to be very reflective on its own role, methods, and impacts.

## Comparative Conclusions

---

Throughout the years, various scholars have pleaded for RA and TA to embrace a broader perspective on risk and/or their object of study (see ➤ *Table 43.2*). Both RA and TA are advised

**Table 43.2****Comparing interpretation of risk within RA and TA**

Science	Risk assessment	Technology assessment
Classical interpretation of risk	Risk = probability × effect Calculable risk based on past experience	(Potential) social effects
Modern interpretation of risk	Calculable risks based on past experience and uncertain (potential) risks based on assessment of the future	(Potential) effects and current social drivers
	• Bad and good chances	• Bad and good chances
	• Role for public perceptions	• New types of effects (e.g., change in our moral vocabulary)
	• Role for broader set of social values to come into play	• Social visions and values that shape technology
	• Reflection on risk governance	• Reflection on role and impact of TA

to look at both bad and good chances. It is suggested that risk assessment should include a broader set of social issues, just like classical TA. Moreover, classical risk approach assumed that risks were calculable based on past experience. Modern risk approaches are increasingly oriented toward the future and should be able to deal with both calculable risks and uncertain (potential) risks. Interestingly, TA is encouraged to broaden its view in the other direction; by mapping not only (potential) effects, but also analyzing current visions and values that drive science and technology. The various shortcomings of RA and TA also constantly provoke discussions about the (in)capacity of government and political institutions to steer technological change from a societal point of view.

## Risk Characteristics and RA and TA

This section describes how, driven by the concept of uncertainty, attempts have been made to characterize risk or problem situations in order to clarify the limitations of the classical approach and to promote and legitimize participatory approaches in the field of risk management and TA.

## Risk Situations and Risk Management Strategies

Inspired by the academic considerations on risk and uncertainty that were described in the former section, also the awareness among risk assessors and managers has grown that different situations of risk require different risk management strategies. In particular, the *White Paper on Risk Governance* (Renn 2005) by the International Risk Governance Council has played a major role in bringing academic ideas into the risk assessment and management field (see also [Chap. 44, Risk Governance](#) by Hermans, Fox, and van Asselt in the present book). Based

**Table 43.3**

Risk characteristics and their implications for risk management (Source: Renn 2005, p. 16)

Knowledge characterization	Risk management strategy	Stakeholder participation
“Simple” risk problem	Routine-based	Instrumental discourse
Complexity-induced risk problems	Risk-informed and Robustness-focused	Epistemological discourse
Uncertainty-induced risk problems	Precaution-based and resilience-focused	Reflective discourse
Ambiguity-induced risk problems	Discourse-based	Participative discourse

on the different states of knowledge about each particular risk, the risk governance framework distinguishes between “simple,” “complex,” “uncertain,” and “ambiguous” risk problems. The argument is that the classical risk approach only suffices for simple risk problems. The other three types of risk situations require other risk strategies.

Resolving complex risk issues requires discussion among experts. Klinke and Renn (2002) plea for an “epistemic discourse,” within which experts argue over the factual assessment and the best estimation for characterizing the risks under consideration. The management of risks characterized by high uncertainties should be guided by a “reflective discourse.” Such a discourse includes policy makers, stakeholder groups, and scientists. Besides dealing with the clarification of knowledge, reflective discourse is about finding a balance between over- and under-protection. Finally, ambiguity-induced risk problems are typified by the fact that risk information is interpreted differently by various stakeholders in society and (potential) intense conflict over values and priorities of what should be protected. According to the International Risk Governance Council (IRGC), this type of risk problems which are characterized by interpretative and normative ambiguity demand a participative discourse. Participative discourses are meant to search for solutions that are compatible with interests and values of the people affected and to resolve conflicts among them (● [Table 43.3](#)).

## Characterizing Problem Situations and Type of TA

Interestingly, also with regards to TA, attempts have been made to clarify the role of TA depending on the problem situation. Even more so, the approach taken by the IRGC has its roots within the practice and theory of participatory TA (cf. Renn 1999). Here, we present a taxonomy developed by Grin et al. (1997) to decide about the role of TA with regards to a certain problem situation (see ● [Table 43.4](#)). According to these authors, there will be little need for TA when there is little uncertainty regarding facts and little value dissent (so-called structured problems). When there is little value dissent, but high uncertainty about the facts (so-called moderately structured scientific problems), the role of TA may be to clarify the facts and their relationships. In such a situation, classical TA suffices. When the problem situation is characterized by a great deal of value dissent, expert-driven analysis is no longer sufficient. In case of (moderately structured or unstructured) political problems, participatory forms of TA need to come into play.

**Table 43.4**

Role and type of TA depending on problem situation (Adapted from Grin et al. 1997, p. 21)

	Uncertainty regarding facts: little	Uncertainty regarding facts: much
Value dissent: little	Structured problem	Moderately structured scientific problem
	<i>Little need for TA</i>	<i>Classical TA to clarify facts and their relationships</i>
Value dissent: much	Moderately structured political problem	Unstructured political problem
	<i>Participatory forms of TA might be beneficial</i>	<i>Participatory forms of TA are required</i>

**Table 43.5**

Comparing the relationship between problem characteristics and their implications for risk management and role and type of TA (Adapted from Grin et al. 1997, p. 21 and Renn 2005, p. 16)

Knowledge characterization	Stakeholder participation within risk management strategy	Role and type of TA
“Simple” risk problem	Classical risk approach: instrumental discourse	Little need for TA
Complexity-induced risk problems	Epistemological discourse	Classical TA to clarify facts and their relationships
Uncertainty-induced risk problems	Reflective discourse	Participatory TA might be beneficial
Ambiguity-induced risk problems	Participative discourse	Participatory TA is required

## Comparative Conclusions

The arrival of the notion of uncertainty regarding to science and values has led to a systematic reflection on the limits of the classical approach to TA and risk (see [Table 43.5](#)). This has led to the insight that in the case of “simple” risk problems, the classical risk approach works well. It is interesting to note that in such structured problem situations, there is little need for TA. In other words, the practices and roles of classical RA and TA do not overlap each other. In case of complex, uncertain, and ambiguity-induced risk problems, both practices start to overlap, or better, complement each other. In particular, the latter two problem characteristics challenge TA and risk assessment and management to introduce participatory approaches in the political decision-making process (see section [New \(Participatory\) Approaches to TA and RA](#)). Systematic reflection has – politically and intellectually – legitimized the need to use and experiment with participatory approaches. In fact, we saw that experience and reflection in the field of participatory TA at the end of the 1990s (Renn 1999) has had a marked influence into the field of risk management in the middle of the first decade of this century, leading to the so-called IRGC risk governance model (Renn 2005).

## New (Participatory) Approaches to TA and RA

A second basic complaint about the expert-based classical approaches concerns the fact that they do not sufficiently represent (broader) public concerns about science and technology. To address this shortcoming, new participatory approaches to TA and RA have been proposed and to a certain extent implemented. These approaches aim to involve a broader set of social actors within the political decision-making process on technology-related risks. This section describes and compares these new participatory approaches.

### New (Participatory) Approaches to TA

We saw that classical TA aims to both rationalize public controversy and aims to strengthen representative democracy by empowering the role of parliaments vis-à-vis the government (see ➤ [Table 43.1](#)). In contrast to classical TA, new (participatory) approaches to TA seek to *represent* public controversies over technological change (Joss 2000). As such, they attempt to organize the interface between the (political) decision-making arena and society in a more interactive manner. This participatory approach is also promoted to improve the interface between TA and the parliament in order to strengthen the impact of TA.

### Improving the Interaction Between TA and Politics

The establishment of OTA inspired MPs in various European countries to start discussing the need for parliamentary TA in their own country. During the 1980s, countries like Germany and Great Britain adapted the classical TA model. The institutionalization of parliamentary TA in Europe, however, also introduced new rationales and roles for TA. For example, the French MPs decided to start doing TA themselves, supported by the staff of OPECTS (Office Parlementaire d’Evaluation des Choix Scientifiques et Technologiques). In this way, the French MPs organized the interface between science and politics in a novel way.

Namely, in the classical TA model, MPs are basically informed by technology assessors through written scientific reports. Communication through writing, which is dominant within academic circles, does not match well the dominant oral culture within parliaments. Moreover, the agendas of MPs are often overloaded. This has challenged parliamentary TA institutes to rethink and redesign the way they communicate and interact with MPs (Decker and Ladikas 2004). It was realized that playing a role in the interface between science and politics does not simply mean bringing scientific insights to the parliament. Instead communication implies a two-way process. This realization has led to more participatory forms of communication, in which issue framing and identification of information needs result from interactions between TA practitioners and MPs. In fact, OPECTS exemplifies such an approach. For other parliamentary TA institutes organizing hearings, expert workshops, and Future Panels present ways to stimulate the (direct) involvement of MPs. For a comprehensive overview of European parliamentary TA development, see Vig and Paschen (2000), see also [www.eptanetwork.org](http://www.eptanetwork.org).

## TA as an Interface Between Politics and Society

Parliamentary TA was also given a new and extra political role in countries like Denmark and the Netherlands. Besides playing a role within the interfaces between parliament and science and government, TA was set up to strengthen the interface between the political arena and society. This was an institutional response to social activism in the 1970s. Public demonstrations, notably around nuclear power, put public authorities under pressure, and created a legitimacy crisis of the State. As a result, controversies over technologies were seen as a problem between the government, the parliament *and* the wider public (Van Eijndhoven 1997). Besides scientifically informing the Parliament, TA was also positioned as a more general and “open” process for involving the public in policy dialogues and building societal consensus on issues of technological change. As an attempt to represent public controversies over science and technology, participatory methods were seen as ways to deal with the interface between the (political) decision-making arena and society. Danish and Dutch MPs saw public engagement and deliberation as a legitimate add-on to representative democracy. As a result, besides monitoring and assessing technological development, the Danish Board of Technology’s (DBT) and the Dutch Rathenau Institute got the task to also stimulate public debate.

In order to fulfill this task, the DBT and the Dutch Rathenau Institute started to experiment with participatory methods to involve experts, stakeholders, and citizens in TA. This involvement has taken various forms, including citizens’ panels and juries, scenario workshops, round tables and consensus conferences. An overview of different participatory methods is presented in Joss (1999) and Slocum (2003). These methods have become more widely established over the last two decades (Joss and Bellucci 2002). In particular, at the beginning of this century, concerns about the science-society relationship and calls for public dialogue became part of the mainstream policy discourse in Europe. In the context of nanoscience, the adjective “upstream” entered the existing discourse on public participation (Wilsdon and Willis 2004). Policymakers and the business and science communities wanted to avoid nanotechnology becoming “the next GM.”

Also, participatory TA has been the target of criticism. First of all, participatory TA is as sensitive to framing as classical TA. In particular, the issue of representation – who should participate and what degree of representativeness should they have – presents an enduring challenge. Secondly, the lack of impact of participatory exercises on the political decision-making process is a central source of concern for parliamentary TA.

## New (Participatory) Risk Approaches

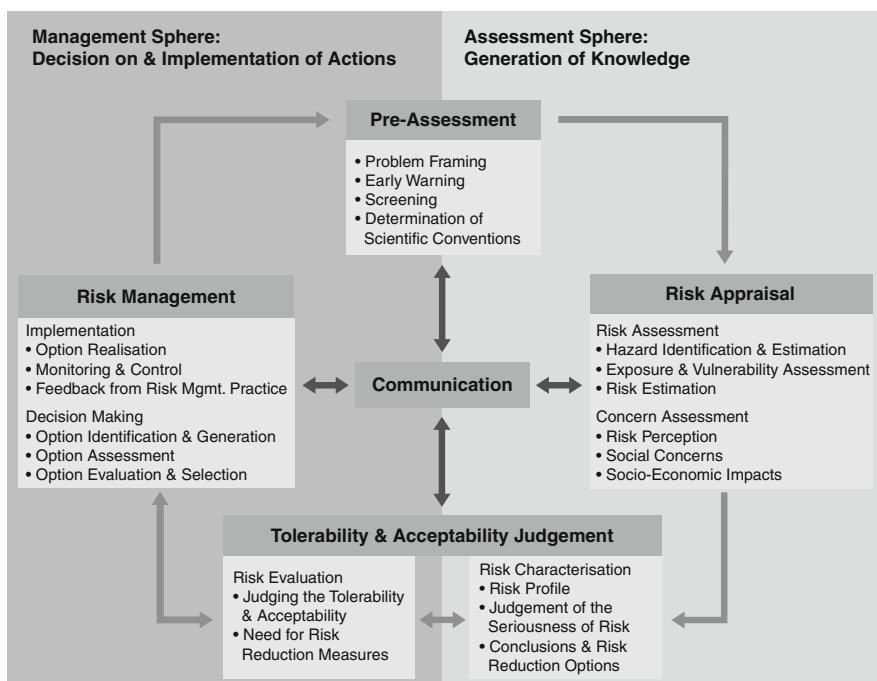
Participation has also been promoted in the field of risk assessment and management (Klinke and Renn 2002; Felt 2007; Klinke 2009). Participation is thought to enable the assessment sphere to take into account broader public concerns on risks. Allowing interested parties to participate in risk appraisal, however, is not yet a common phenomenon. Besides, participation is employed to improve the interface between risk assessment and risk management. One of the main promoters to include deliberation in the risk field is the International Risk Governance Council (IRGC).

## The IRGC Risk Governance Framework

In the wake of BSE and the GMO controversy, a new risk approach developed in the first decade of this century (WRR 2008). The risk governance framework of the International Risk Governance Council (IRGC) forms the most prominent and most elaborated example of this new approach (Renn 2005). As we saw above, the IRGC risk governance framework distinguishes between four types of problem situations and related risk management strategies (see [Fig. 43.3](#)). In particular, in the case of uncertainty-induced and ambiguity-induced risk problems, the IRGC advises to make use of participatory exercises.

### Risk Assessment Plus Concern Assessment

In the classical risk approach risk assessors are assumed to be able to calculate risks, and in that way would present the public interest. This assumption about the science – society interface has been severely criticized. In the risk governance framework, risk assessment is complemented by concern assessment (see [Fig. 43.1](#)) (Renn 2005, pp. 12–15). Together they form the risk appraisal phase. Concern assessment is about getting “knowledge of stakeholders’ concerns and questions – emotions, hopes, fears, apprehensions – about the risk as well as likely social consequences, economic implications, and political responses” (Renn 2005, p. 14).



**Fig. 43.1**

IRGC risk governance framework. (Source: Renn 2005, p. 13, with kind permission from the International Risk Governance Council)

## Improving the Interface Between the Risk Assessment and Management Sphere

The classical risk approach is based on a clear distinction of risk assessment and risk management. In this model, the usability of the results of the expert-based risk assessment for policy decision making are taken for granted. The IRGC risk governance framework includes various extra activities in the risk-handling chain to improve the interaction between the risk assessment sphere and the risk management sphere (see Fig. 43.1) (Renn 2005, pp. 12–15). The framework proposes to include a pre-assessment phase. Risk framing is the first step of this stage. This exercise presents an attempt to come to a common understanding among relevant social actors on what kind of risk issues should be addressed.

After the risk appraisal phase, the risk governance framework proposes to have the stage of risk characterization and evaluation. Risk characterization is a decision-driven activity, directed toward informing decision makers, and is, therefore, the final and most controversial part of the risk assessment sphere. For long, risk characterization was commonly seen as a summarization of scientific information (National Research Council 1983, p. 20). In the early 1990s, the National Research Council in the United States started to promote risk characterization as a process which combines analysis and deliberation (Stern and Fineberg 1994).

## Comparative Conclusions

Participatory approaches have been promoted and developed both in parliamentary TA and RA practices. These participatory exercises try to deal with the interfaces between TA and RA and society and the (political) decision-making arena (see Table 43.6). While classical TA aims to rationalize public controversy, participatory TA seeks to represent public controversy. For RA, something similar is the case. Within the risk governance model, the risk assessment

**Table 43.6**  
Comparing the new approach to risk and parliamentary TA

	New risk approach	New parliamentary TA
Science	Risk appraisal = risk + concern assessment	Technology assessment
Modern interpretation of risk	Calculable risks based on past experience and uncertain (potential) risks based on assessment of the future	(Potential) effects and current social drivers
Political role		
Interface science – policy making	Informing risk managers within the government about technical risks	Informing MPs on social effects of S&T
	Risk characterization based on analysis and deliberation	Involving MPs in TA projects
Interface society – policy making	Pre-assessment: problem framing and early warning	Participatory parliamentary TA: representing public controversies over technological change
	Concern assessment	

sphere is complemented with new types of analysis and deliberation – notably, concern assessment, risk framing, and risk characterization (see [Fig. 43.1](#)) – that aim to include the concerns of various social actors. Besides improving the interface between society and RA or TA, participatory approaches can also be used to improve the interface between TA and MPs and RA and risk managers.

If we take the risk governance framework as the exemplar of the new risk approach, it becomes clear that from a conceptual point of view TA and RA have grown closer to each other over the last decade. Actually, concern assessment, risk framing, and risk characterization, which combine analysis and deliberation, can be seen as typical TA-like activities. In fact, from the perspective of risk governance parliamentary TA and risk assessment and management all have an important role to play. Unfortunately, the complementary roles of these practices have not yet been given enough academic consideration. The next section explores the potential synergy between parliamentary TA and RA in the risk governance of nanotechnology in the Netherlands.

## An Example: Risk Governance of Nanotechnology in the Netherlands

---

In 2003, the public debate on nanotechnology was in its infancy. At the same time, a strong growth in worldwide funding and patenting combined with high scientific and social hopes and concerns fueled the debate. The Dutch government invested large sums in nanoscience, mainly within the framework of the national research program NanoNed. The debate in Europe received a strong impulse, when the ETC Group, an international civil society organization, called for attention to the issue of nanotoxicity at the European Parliament. Triggered by these events, the Rathenau Institute, the Dutch parliamentary TA organization, started activities which aimed to stimulate the public and political debate on the social and ethical issues related to nanoscience and -technology. This section presents some examples of various political roles played by parliamentary TA within the risk governance of nanotechnology.

### Setting the Agenda

---

The Rathenau Institute started by conducting a TA study, to provide an initial concept agenda for public and political discussion (Van Est et al. [2004](#)). That first agenda included a broad range of topics, ranging from health effects of nanomaterials to privacy issues related to nanoelectronics (and the vision of smart environments), and ethical issues related to human enhancement. The Rathenau Institute also wrote a position paper (Van Keulen and Van Est [2004](#)) and organized an expert-stakeholder workshop in February 2004 on the chances and risks of nanomaterials. At the workshop, the nanoscientific community was confronted with societal and policy actors for the first time. As a result, the government commissioned exploratory studies by the National Institute for Public Health and Environment (RIVM) (Roszek et al. [2005](#)) and the Health Council (Gezondheidsraad [2006](#)). Researchers at RIVM had wanted to study the safety aspects of nanomaterials before, but had lacked funding because the issue had not been on the policy agenda before. These activities of the Rathenau Institute helped in overcoming that deadlock and got a wider risk governance process going.

In 2004, the Rathenau Institute continued by writing position papers and organizing workshops on various application fields of nanotechnology. All these public activities fed into a parliamentary hearing at the end of that year. Besides creating a first agenda for discussing nanotechnology, the Institute stimulated the development of an initial heterogeneous network of actors – consisting of nanoscientists, policy makers, politicians, social scientists, and people from industry and civil society organizations – to become involved in the public debate. In response to various discussions with the parliament, the government announced at the end of 2005 to come up with an integral policy on nanotechnology. An interdepartmental working group for nanotechnology (ION) was set up to prepare such a vision.

## Setting Up the Risk Governance Machinery

---

In the political discussion on dealing with the risks of nanotechnology, the advice of the Dutch Health Council (Gezondheidsraad 2006) played an important role. The Health Council produced an extensive report, which used the IRGC risk governance framework (Renn 2005) as a model for dealing in an integral manner with complex, uncertain and even ambiguous risks related to nanotechnology. Anticipating the political significance of the Health Council advice, the Rathenau Institute organized a workshop which brought together key experts and stakeholders to discuss the implications of the Health Council advice for priorities in risk policy (Malsch 2006). The workshop was attended by ION members who were preparing the Cabinet View on Nanotechnologies (Kabinet 2006).

In the Netherlands extensive discussions on dealing with systemic risks had led to the development of a particular risk governance framework, called *Dealing sensibly with risks* (VROM 2004). Within this model, key strategies for dealing with risk are transparent decision making, a clear distribution of responsibilities, early involvement of citizens, balancing risks and benefits, and factoring in accumulation of risks. The Cabinet view was based on this model. In addition, however, the government announced to identify ethical and social issues relating to nanotechnology developments and to set up a broad public dialogue on these issues. From this point on, the Rathenau Institute focused on three challenges for risk governance: dealing with possible physical risks, how to organize a public dialogue, and further developing the wider public agenda, in particular, related to technological convergence (Van Est and Walhout 2010). We will focus here on the first two issues.

## Dealing with “Organized Irresponsibility”

---

With regards to dealing with nanotoxicity, the government followed the European Commission’s statement that the existing regulatory frameworks are sufficient to deal with nanomaterials, apart from specific amendments to be made. These regulations assign responsibility for safety primarily to producers. The government also recognized its responsibility for setting up the participatory aspect of risk governance. The government asked the Social Economic Council – an established negotiating platform for employer organizations and labor unions – to come up with an advice on occupational safety. Besides, a broad sounding

board on nanosafety became part of the interactive policymaking the government already deployed for consumer and environmental affairs.

Assigning responsibilities and facilitating dialogue does not automatically leads to a proactive handling of nanotoxicity. By closely following the risk discussion, the Rathenau Institute recognized that the actual risk governance of nanotoxicity was hampered by a number of mutual dependencies, which caused acting voids and deadlocks. For example, the combination of a lack of knowledge, definitions, and oversight caused a lack of funding and inability for priority setting. This results into a lack of sense of urgency, which maintains the current deadlock. The Rathenau Institute stimulated a political discussion about these examples of “organized irresponsibility” (Beck 1992). Together with a parliamentary committee, it organized in 2009 a parliamentary hearing, which allowed MPs to hear the voice of a broad range of experts and stakeholders. The Institute also advised the parliament about priorities for the political discussion.

## Monitoring and Debating Risk Governance

---

In line with the suggestions of the Health Council, and thus the IRGC risk governance framework, the government had decided to organize a public dialogue on social and ethical issues. According to that model, uncertainty about physical risks was to be discussed in expert-stakeholder settings, and more ambiguity-induced risk problems required communication and social discourse (see section [❸ Risk Characteristics and RA and TA](#)). In practice, however, these issue categories and the related deliberative processes cannot be separated. In other words, these governance processes overlap and relate to each other. This has to do with the fact that both discussions often involve similar types of social actors, and with the priorities these actors have. When the issue of physical risks is surrounded with many uncertainties, most people are inclined to focus on such a concrete issue and tend to ignore or postpone a debate on often more intangible social and ethical questions. The Rathenau Institute highlighted this condition in the publication *Ten lessons for a nanodialogue* (Hanssen et al. 2008). Moreover, it advised the government about do’s and don’ts with regard to setting up a public dialogue on nanotechnology.

## Parliamentary TA’s Role in Risk Governance

---

The Dutch nanotechnology case shows that parliamentary TA can play various political roles within the risk governance process. Parliamentary TA may play a role in developing the public and political agenda. Parliamentary TA developed an initial public agenda and was instrumental in putting nanotechnology on the agenda of the government and the parliament. By using all kinds of participatory methods, parliamentary TA has also stimulated public and political debate by involving all kinds of stakeholders. Finally, parliamentary TA stimulated critical reflection and debate on the risk governance approach and processes itself. With regards to the safety issue, the Rathenau Institute has exposed various regulatory uncertainties that prevent effective governance. Moreover, it clarified that whereas the IRGC risk governance proposes different risk strategies for different risk issues, the practice of risk governance often does not allow for such a clear separation of issues.

## Further Research

---

The structure of this chapter provides an interesting agenda for future research.

We started this chapter by stating that the relationship between risk and technology assessment has been scarcely discussed. This chapter shows that such an exercise is fruitful since parliamentary TA may form a source of inspiration for RA. Whereas RA originally used a very narrow definition of risk, TA employed a common sense view on risk. The latter is due to the fact that parliamentary TA, because of its mission and institutional position, is continuously confronted with the political arena and society. As such, it has to deal with analysis and deliberation on the one hand and power on the other. Defining itself as “science,” the risk field has (had) a hard time in dealing with the two poles of political practice. A crucial step is accepting a broader interpretation of risk, nowadays captured by the term “uncertain risk.” TA’s history, however, shows yet another challenge. Its focus on social effects was broadened with a view on social drivers. This links to the argument that RA and TA should pay equal attention to both “bad” and “good chances” (Wildavsky 1988), and the recent plea for a “philosophical hope/risk analysis” (Ferrari and Nordmann 2009). How to deal with such a broad interpretation of risk and the positive side of risk taking presents an interesting theme for further research.

A second research theme concerns the relationship between risk characteristics and the role and type of TA or the appropriate risk management strategy. Scholars like Renn (1999, 2005) and Grin et al. (1997) present their taxonomies as conceptual guidelines, instead of rigid normative prescriptions. Still these categorizations play an important normative and political role in the discourse on RA and TA. It would be relevant to study this, but also to analyze how the relationship between risk characteristics and the choice for a certain risk management strategy or TA approach is shaped in practice. In this respect, Van Asselt et al. (2009) talk of the so-called “uncertainty paradox.” While there is great awareness among risk managers that a lot of uncertainties are in play, still there is a strong (institutional) tendency to use science to diminish uncertainty. These types of mechanisms are in need of more reflection and analysis.

Related to this, it would be relevant to map in what kind of situations and to what extent new (participatory) approaches to TA and RA are actually used. Do these type of deliberative exercises still play a marginal role or have they become fully integrated? As we saw also participatory approaches face problems with regards to problem framing, representation, timing and impact. There is a need for more insight into how risk and technology assessors view and deal with these issues in practice. Moreover, modern approaches to TA and RA (in particular, the IRGC risk governance framework) are about finding a balance between analysis and deliberation. There is still little knowledge on how science and involvement of various social actors is organized in practice. In particular, there seems to be a large gap between academic research done in this field and the way practitioners view and reflect on their practices.

This brings us to the big theme of risk governance. The history of TA and RA shows numerous attempts to deal with the social risks involved in science and technology. They present institutional attempts to address the democratic deficits of the modernist practice of managing technology. In *Risk Society*, Beck (1992) hinted at the above central research theme by dropping the term “organized irresponsibility.” The IRGC risk governance framework is a great attempt to develop an integral vision on risk governance. Namely, besides “risk assessment” and “risk management,” this framework also aspires to look at “how risk-related

decision-making unfolds when a range of actors is involved, requiring coordination and possibly reconciliation between a profusion of roles, perspectives, goals and activities" (Renn 2005, p. 11). In this chapter, we have indicated that within the classical approach to risk and TA, the gap between RA and TA is impossible to narrow. In contrast, the new risk governance approach, in principle, has the potential to bridge these two practices. To put it stronger: parliamentary TA has a clear role to play within risk governance. Up to now, however, the democratic role parliamentary TA currently plays and may play within risk governance has received very little attention. To conclude: parliamentary TA's role within risk governance presents a remarkable blind spot on the current research agenda.

## References

- Beck U (1992) Risk society: towards a new modernity. Sage, London
- Bereano P (1997) Reflections of a participant-observer: the technocratic/democratic contradiction in the practice of technology assessment. *Technol Forecast Soc* 54(2&3):163–175
- Bimber B, Guston DH (1997) Technology assessment – the end of OTA. *Technol Forecast Soc* 54(2&3):125–308
- Coates JF (2001) A 21st century agenda for technology assessment. *Technol Forecast Soc* 67:303–308
- Decker M, Ladikas M (eds) (2004) Bridges between science, society and policy. Technology assessment – methods and impacts. Springer, Berlin/Heidelberg/New York
- Everson M, Vos E (eds) (2009) Uncertain risks regulated. Routledge-Cavendish, New York
- Felt U (rapporteur) (2007) Taking European knowledge society seriously. Report of the expert group on science and governance. European Commission Science, Directorate-General for Research, Directorate Science, Economy and Society, Brussels
- Ferrari A, Nordmann A (eds) (2009) Reconfiguring responsibility: lessons for nanoethics (Part 2 of the report on deepening debate on nanotechnology). Durham University, Durham
- Fischer F, Forester J (eds) (1993) The argumentative turn in policy analysis and planning. Duke University Press, Durham
- Fischhoff B, Lichtenstein S, Slovic P, Derby SL, Keeney RL (1981) Acceptable risk. Cambridge University Press, Cambridge
- Gezondheidsraad (2006) Betekenis van nanotechnologieën voor de gezondheid. Gezondheidsraad, Den Haag
- Grin J, Van de Graaf H, Hoppe R (1997) Technology assessment through interaction: a guide. Rathenau Institute, The Hague
- Grin J, Grunwald A (eds) (2000) Vision assessment: shaping technology in the 21st century. Towards a repertoire for technology assessment. Springer, Berlin/Heidelberg/New York
- Grunwald A (2009) Technology assessment: concepts and methods. In: Dov M, Meijers A, Woods J (eds) Handbook of the philosophy of science. Philosophy of technology and engineering sciences, vol 9. Elsevier, Amsterdam, pp 1103–1146
- Hanssen L, Van Walhout B, Est R (2008) Ten lessons for a nanodialogue: the Dutch debate about nanotechnology thus far. Rathenau Institute, The Hague
- Harremoës P, Gee D, MacGarwin M, Stirling A, Keys J, Wynne B, Guedes Vaz S (2001) Late lessons from early warnings: the precautionary principle 1896–2000. European Environment Agency, Copenhagen
- Hood C, Rothstein H, Baldwin R (2001) The government of risk: understanding risk regulation regimes. Oxford University Press, Oxford
- Jaspers K (1965) Kleine Schule des philosophischen Denkens. R.Piper, München
- Joss S (ed.) (1999) Special issue on public participation in science and technology. *Science and Public Policy* 26(5):289–380
- Joss S (2000) Participation in parliamentary technology assessment: from theory to practice. In: Vig NJ, Paschen H (eds) Parliaments and technology: the development of technology assessment in Europe. State University of New York Press, New York, pp 325–364
- Joss S, Bellucci S (eds) (2002) Participatory technology assessment: European perspectives. Centre for the Study of Democracy, London
- Kabinet (2006) Cabinet view on nanotechnologies. Kabinet, Den Haag
- Klinke A, Renn O (2002) A new approach to risk evaluation and management: risk-based, precaution-based, and discourse-based strategies. *Risk Anal* 22(6): 1071–1094
- Klinke A (2009) Inclusive risk governance through discourse, deliberation and participation. In:

- Everson M, Vos E (eds) *Uncertain risks regulated*. Routledge-Cavendish, New York
- Knight FH (1921) *Risk, uncertainty and profit*. Houghton Mifflin, Boston/New York
- Kunkle GC (1995) New challenge or the past revisited? The office of technology assessment in historical context. *Technol Soc* 17(2):175–196
- Lynn LE (1981) *Managing the public's business: the job of the government executive*. Basic Books, New York
- Maasen S, Merz M (2006) TA Swiss broadens its perspective. Technology assessment with a social and cultural sciences orientation. TA Swiss Centre for Technology Assessment, Bern
- Malsch I (2006) *Verslaglegging Expertmeeting milieu- en gezondheidsrisico's van nanodeeltjes: Naar een prudent beleid*. Rathenau Instituut, Den Haag
- National Research Council (1983) *Risk assessment in the federal government: managing the process*. National Academy Press, Washington, DC
- Paschen H (2000) The technology assessment bureau of the German parliament. In: Vig NJ, Paschen H (eds) *Parliaments and technology: The development of technology assessment in Europe*. State University of New York Press, New York, pp 93–124
- Perrow C (1984/1999) *Normal accidents: living with high-risk technologies*. Princeton University Press, Princeton
- Renn O (1999) Participative technology assessment: meeting the challenges of uncertainty and ambivalence. *Futures Res Q* 15(3):81–97
- Renn O (2005) IRGC White Paper No. 1: risk governance – towards an integrative approach. International Risk Governance Council (IRGC), Geneva
- Roszek B, De Jong WH, Geertsma RE (2005) Nanotechnology in medical applications: state of the art in materials and devices. RIVM, Bilthoven
- Slocum N (2003) *Participatory methods toolkit. A practitioner's manual*. King Baudouin Foundation and Flemish Institute for Science and Technology Assessment (viWTA), Brussels
- Slovic P (1978/2000) *The perception of risk*. Earthscan, London
- Stern PC, Fineberg HV (eds) (1994) *Understanding risk: informing decisions in a democratic society*. National Academic Press, Washington, DC
- Swierstra T, Van Est R, Boenink M (2009) Taking care of the symbolic order: how converging technologies challenge our concepts. *J Nanoethics* 3(3):269–280
- U.S. Congress (1972) *The Technology Act of 1972*. Public Law 92–484, H.R. 10243, 13 Oct 1972
- Van Asselt M, Vos E, Rooijakkers B (2009) Science, knowledge and uncertainty in EU risk regulation. In: Everson M, Vos E (eds) *Uncertain risks regulated*. Routledge-Cavendish, New York
- Van Eijndhoven J (1997) Technology assessment: product or process? *Technol Forecast Soc* 54(1997): 269–286
- Van Est R, Malsch I, Rip A (2004) *Om het kleine te waarderen: Een schets van nanotechnologie: Publiek debat, toepassingsgebieden en maatschappelijke aandachtspunten*. Rathenau Instituut, Den Haag
- Van Est R, Walhout B (2010) Waiting for nano – very actively: a long-term view on the role of the Rathenau Institute in stimulating the Dutch debate on nanotechnology. *Technikfolgenabschätzung – Theorie und Praxis* 2 (July) 67–74
- Van Est R, Brom F (2012). Technology assessment as an analytic and democratic practice. In: *Encyclopedia of applied ethics*. (2nd edition) Elsevier Science, Amsterdam
- Van Keulen I, Van Est R (2004) *Gezondheids- en milieurisico's van nanodeeltjes: achtergrondinformatie voor de Themapcommissie Technologieberaad*. Rathenau Instituut, Den Haag
- Vig NJ, Paschen H (eds) (2000) *Parliaments and technology: the development of technology assessment in Europe*. State University of New York Press, New York
- VROM (2004) *Nuchter omgaan met risico's: Beslissen met gevoel voor onzekerheden*. Ministerie van VROM, Den Haag
- Wildavsky A (1988) *Searching for safety*. Transaction, New Brunswick
- Wilsdon J, Willis R (2004) *See-through science: why public engagement needs to move upstream*. Demos, London
- WRR – Wetenschappelijke Raad voor het Regeringsbeleid (2008) *Onzekere veiligheid: Verantwoordelijkheden rond fysieke veiligheid*. Amsterdam University Press, Amsterdam



# 44 Risk Governance

Marijke A. Hermans · Tessa Fox · Marjolein B. A. van Asselt

Faculty of Arts & Social Sciences, Maastricht University, Maastricht,  
The Netherlands

<b>Introduction .....</b>	<b>1094</b>
<b>The Origins of Risk Governance .....</b>	<b>1095</b>
Risk .....	1095
Positivistic Risk Paradigm .....	1096
Enlightened Engineering and Psychology .....	1096
Anthropology/Sociology .....	1097
Social Amplification of Risk Framework (SARF) .....	1098
Science and Technology Studies (STS) .....	1098
Political Sciences .....	1099
Law .....	1099
Reflection .....	1100
Governance .....	1101
The Origins of Risk Governance .....	1102
The Typology of Risk .....	1104
<b>Risk Governance .....</b>	<b>1108</b>
Risk Governance Principles .....	1108
The Communication and Inclusion Principle .....	1109
The Integration Principle .....	1110
The Reflection Principle .....	1111
<b>Conclusion and Discussion .....</b>	<b>1112</b>

**Abstract:** Recently, the notion of risk governance has been introduced in risk theory. This chapter aims to unravel this new concept by exploring its genesis and analytical scope. We understand the term “risk governance” as the various ways in which many actors, individuals, and institutions – public and private – deal with risks surrounded by uncertainty, complexity and/or ambiguity. It includes, but also extends beyond, the three conventionally recognized elements of risk analysis (risk assessment, risk management, and risk communication). Risk governance emphasizes that not all risks can be calculated as a function of probability and effect. We argue that risk governance is more than only the critical study of complex, interacting networks in which choices and decisions are made around risks; it should also be understood as a set of normative principles which can inform all relevant actors of society on how to deal responsibly with risks. In this chapter, we take stock of the current body of scholarly ideas and proposals on the governance of contemporary risks along the lines of three principles: the communication and inclusion principle, the integration principle, and the reflection principle.

## Introduction

---

Over the past few decades, modern society has been increasingly challenged to manage potentially negative outcomes of technological developments. The nature of these hazards, as well as their lack of temporal and spatial limits, has given rise to a call for a new integrative approach that aims to understand, assess, and handle risks to human health and the environment by building upon and extending current risk analysis practices. Many discussions about technological innovation occur nowadays in the public arena, which concern not only health and safety but also ethical and social issues. Arguably a decline of public trust in the ability of experts and policy makers to deal with risks has been accompanied by a growing demand for public participation in scientific and technical decision making (Adams 2005; Beck 1992; Giddens 1991; Wynne 1982).

How can we deal with these increasingly complex and demanding risks and with the need for more public involvement? One set of proposals is clustered under the heading “risk governance,” a combination of the terms “governance” and “risk.” “Risk governance” aims to provide an approach for how to deal responsibly with public risks. Risk governance pertains to the various ways in which many actors, individuals, and institutions – public and private – deal with risks surrounded by uncertainty, complexity and/or ambiguity. It includes, but also extends beyond, the three conventionally recognized elements of risk analysis: risk assessment, risk management, and risk communication. It moreover requires consideration of the legal, institutional, social, and economic contexts in which risk is evaluated, as well as consideration of the interests and perspectives of different actors and stakeholders. In sum, risk governance aims to take into account the complex web of actors, rules, conventions, processes, and mechanisms concerned with how relevant risk information is collected, analyzed, and communicated, and how management decisions are taken (IRGC (International Risk Governance Council) 2005, 2007; Renn 2008; Renn et al. 2011; van Asselt and Renn 2011; van Asselt and Vos 2008).

Risk governance has its origins in the scholarly ideas on how to deal with demanding public risks informed by several decades of interdisciplinary research drawing from engineering studies, psychological and sociological research on risk, science and technology studies (STS), social movement theory, and research by policy scientists and legal scholars, including Ravetz (e.g., Funtowicz and Ravetz 1992; Ravetz 1996 [1971]), Nowotny (e.g., Nowotny 1976, 2008; Nowotny et al. 2001), Fischhoff (e.g., Fischhoff et al. 1984), Slovic (e.g., Slovic 1987, 2000),

Wynne (e.g., Irwin and Wynne 1996; Wynne 1982, 2002a), O’Riordan (e.g., O’Riordan 1982; O’Riordan et al. 2001; O’Riordan and McMichael 2002), Beck (e.g., Beck 1992, 2009; Beck et al. 1994), Jasenoff (e.g., Jasenoff 1990, 2005), Tesh (Tesh 2000), the Kaspersons (e.g., Kasperson and Kasperson 1991, 2005a, 2005b, see also Pidgeon et al. 2003), Löfstedt (e.g., Löfstedt 2005; Löfstedt and Renn 1997), Stirling (e.g., Stirling 1998, 2004), van Asselt and Vos (e.g., van Asselt 2000; van Asselt and Vos 2006, 2008; Vos 2000, see also Fisher 2008), Fischer (e.g., Fischer 2002; Fischer et al. 2006), Huitema (Huitema 2002), Mourik (Mourik 2004), and so-called cultural theorists (e.g., Adams 2005; Douglas and Wildavsky 1982; Rayner 1992; Thompson et al. 1990). This body of knowledge provides a convincing and empirically sound basis to argue that many risks cannot be calculated on the basis of quantitative methods alone, and that regulatory models which build on this assumption are not just inadequate, but form a complication to responsibly dealing with many contemporary risks.

In this chapter, we will explain the origins of risk governance and the issues and proposals that fall under this heading. Is risk governance indeed a major change in the ways in which risks are conceptualized, appraised, regulated, and communicated as proponents claim? The contours of this framework are derived from the work of a number of prominent scholars, whom we will discuss in this chapter, but the process of turning these empirically informed theoretical proposals into practical reality is still in its infancy. Hence, this chapter aims to provide an original perspective on risk theory since it examines what is referred to as a “paradigm shift” compared to the classic, quantitative, probability-based approach to risk assessment, management, and communication.

We will first describe the origins of the concept (section [The Origins of Risk Governance](#)), reflecting on the most important scientific approaches and disciplines in risk research that have contributed to the emergence of risk governance (section [Risk](#)). Next, we discuss risk governance as part of the broader governance turn in policy sciences (section [Governance](#)), and its context of origin (section [The Origins of Risk Governance](#)). Following the state-of-the-art review by van Asselt and Renn (Renn et al. 2011; van Asselt and Renn 2011), we will discuss what is agreed to be needed to address uncertain, complex, and ambiguous risks adequately (section [The Typology of Risk](#)). The various ideas and proposals pertaining to risk governance are discussed in terms of a set of principles: the communication and inclusion principle, the integration principle, and the reflection principle (section [Risk Governance](#)). These principles aim to synthesize the most important aspects in organizing structures and processes to govern risks.

## The Origins of Risk Governance

---

### Risk

---

The urge to suppress and control risks has been a human endeavor since the ancient Greeks, followed in modern times by the prominent idea that risks are manageable and measurable (Bernstein 1996). This positivistic, quantitative approach to risk, in which estimation of probability and effect is central, has been and still is the dominant way of conceptualizing, assessing, and managing risks. Only recently an academic change of focus has emerged realizing that many contemporary risks reach beyond this traditional definition of risk as calculable and predictable. The social experience of risk is not confined to or not even expressible as

a technical definition of risk, namely, the product of probability and effect. The way people conceptualize and deal with risks in their everyday lives is influenced by values, attitudes, social influences, and cultural identity. We will describe these critical developments in different strands of literature that have contributed to the current understanding and thus explicitly or implicitly to the idea of risk governance. These contributions are very multidisciplinary, ranging from psychology and law to science and technology studies (STS).

Summarizing the work of many authors for several decades cannot be done without compromising historical or scholarly accuracy. However, this overview aims to sketch a general picture of the broad sweep of major ideas and empirically informed insights that have given rise to risk governance. We therefore do not claim to be exhaustive.

## Positivistic Risk Paradigm

In the beginning of the twentieth century, two influential economists delved into the problems of risk. Frank Knight published *Risk, Uncertainty and Profit* (Knight 1921) in the same year as John Maynard Keynes published *A Treatise on Probability* (Maynard Keynes 1921 [2004]). Both scholars addressed the problems of making choices in uncertain circumstances and both defined the concept of “risk” and “uncertainty,” albeit in opposite ways.

Knight argued that it is possible and necessary to sharply distinguish risk from uncertainty: *risk* can be explained as “you don’t know for sure what will happen, but you know the odds,” while *uncertainty* means that “you don’t even know the odds” (Adams 2005). Thus, Knightian uncertainty is immeasurable and not calculable. By contrast, risk is measurable by using the formula:  $\text{risk} = \text{chance} \times \text{effect}$ . Keynes, on the other hand, did not distinguish risk from uncertainty. He claimed that life was dominated by uncertainty, not probability. If life would obey to the laws of probability, humans would have no choices and no influence on the course of events. He therefore stressed the positive associations of uncertainty. In a way, as will become clearer throughout his chapter, risk governance could be viewed as inheriting from the Keynesian view on uncertainty and risk.

However, the Knightian definition dominated and is still dominant in practices of risk assessment and management and in many scientific disciplines such as engineering and economics, which use technical risk analyses to calculate expected benefits and monetary costs, or as part of engineering planning and design. Knight’s concept of risk narrows the focus to the probability of events and the magnitude of specific consequences. It has had a major influence on risk regulation, becoming the “golden formula” and the basis of what is referred to as the classical or *positivistic risk approach*, which developed in the course of the twentieth century. Following the classical risk approach, two phases are distinguished and are ideally institutionally separated: identification and evaluation of risk (risk assessment) and taking measures to control risks that are deemed unacceptable (risk management).

Below we will discuss research traditions and contributions that implicitly or explicitly relate to a more Keynesian view on uncertainty and risk.

## Enlightened Engineering and Psychology

From the 1960s onward, we notice an important turn in risk research from a narrow positivistic risk focus to an approach incorporating qualitative, social, cultural, and normative aspects that

are believed to be an intrinsic aspect of complex risks. In the 1960s, some engineers expanded their risk research from a pure technical exercise into a focus on risk perception, namely on how individuals perceive risks in everyday life. Starr (1969) developed a quantitative method to look at how people weigh risks and advantages. He was one of the first to show that people accept activities that are voluntary (e.g., smoking) more than those that are involuntary (e.g., living next to nuclear power plant). He furthermore introduced the distinction between “objective” and “perceived” risk, to discriminate between the “scientific definition” and the “lay perception” (Starr and Whipple 1980). Risk perception research, taken up by cognitive psychologists after Starr’s pioneering work, built upon Starr’s observation that experts and the public often have different notions of what constitutes a risk. This observation became particularly poignant in the fierce public resistance against nuclear energy in the 1960s and 1970s. Even though scientific experts declared nuclear energy as a safe and clean form of energy, the public perceived it as a threat and protested against it.

The field of psychology has had a major influence on risk studies since questions arose about the publics’ acceptance of risks. Psychologists claim that the public acceptance and reluctance to take risks needs to be explored in relation to the complexities of the human mind (Bouder 2008). Psychologists have dramatically increased our understanding of individual risk decisions and their wider impact on society. Challenging Starr’s strict voluntary/involuntary model, so-called *psychometric studies*, rooted in psychology and decision theory, focused on the roles of affect and emotions in influencing risk perception. Key characteristics such as familiarity, control, catastrophic potential, equity, and level of knowledge have been proven to influence risk decisions (Slovic 2000; Slovic et al. 1982; see also van Asselt 2000 for an overview). The psychometric paradigm, using standardized questionnaires and large-scale surveys, has become one of the dominant approaches in the field of risk research.

## Anthropology/Sociology

In the 1980s, anthropologist Mary Douglas and political scientist Aaron Wildavsky moved beyond the focus on the individual and his/her subjective estimates and challenged the dominance of the psychometric paradigm by publishing *Risk and Culture* (Douglas and Wildavsky 1982) in which they introduced the “Cultural Theory of Risk.” Cultural Theory stresses the importance of culture and society in shaping perceptions of risk. Cultural theorists analyze social responses to risk as being determined by cultural belief patterns that encourage individuals and social groups to adopt certain values and reject others. Cultural Theory outlines four “ways of life” in a group/grid typology: fatalism, hierarchy, individualism, and egalitarianism. By defining risk as a socially constructed phenomenon, Douglas and Wildavsky aimed to show the limitations of quantitative risk assessment that pin down risk in objective measurements and the limitations of psychometric studies that neglect social and cultural influences on risk perception. They challenged the objective-perceived dichotomy:

- ▶ The main questions posed by the current controversies over risk show the inappropriateness of dividing the problem between objectively calculated physical risks and subjectively biased individual perceptions. (...) Between private, subjective perception and public, physical science there lies culture, a middle area of shared beliefs and values. The present division of the subject that ignores culture is arbitrary and self-defeating (Douglas and Wildavsky 1982, p. 194).

In the last decades, this approach has been further developed by other authors (Rayner 1992; Schwarz and Thompson 1990; Thompson, et al. 1990; Wildavsky and Dake 1990) and applied in various contexts (see e.g., Rotmans and de Vries 1997). Even though the validity of these prototypical descriptions has been debated, cultural analysis has indicated that there is not one single, universal approach and conceptualization of risk (Renn 1998).

## Social Amplification of Risk Framework (SARF)

In the 1980s, the development of sociological risk research was heavily influenced by fierce public resistance and controversies regarding new technologies such as nuclear power and disasters such as in Bhopal and Chernobyl. Several prominent scholars tried to integrate the research on the public experience of risk from psychology, anthropology, sociology, and communication studies into an interdisciplinary framework called “Social Amplification of Risk Framework” (SARF) (Kasperson et al. 1988; Pidgeon et al. 2003).

- ▶ The framework aims to examine broadly, and in social and historical context, how risk and risk events interact with psychological, social, institutional, and cultural processes in ways that amplify or attenuate risk perceptions and concerns, and thereby shape risk behavior, influence institutional processes, and affect risk consequences (Pidgeon et al. 2003, p. 2).

The main thesis of SARF is that information processes, institutional structures, social behavior, and individual responses shape the social experience of risk in ways that either increase or decrease public perceptions of risk. SARF tries to explain why risks or risk events assessed as minor by experts might produce massive public reactions, and even have substantial social and economic impacts (risk amplification), while other risks that have been assessed by experts as dangerous do not produce anxious reactions, but are almost ignored (risk attenuation). Examples of “risk attenuation” are smoking or traffic accidents. “Risk amplification” can fuel risk controversies, such as around nuclear energy or genetically modified organisms. The framework tries to integrate the technical assessment of risk with the social and cultural experience of risk, while at the same time it is questioned whether it offered additional knowledge for understanding risks (Zinn and Taylor-Gooby 2006).

## Science and Technology Studies (STS)

Scholars in science and technology studies (STS), or more specifically sociology of scientific knowledge (SSK), critically examine the role and place of technology and science in contemporary society in an interdisciplinary way (Hess 1997b). STS is not only the study of how modern societies are constituted by science and technology and how science and technology affect society, politics, and culture, but also how reciprocally, cultural, social, and political factors determine technological and scientific developments. STS cannot be defined as a homogeneous field, but is composed of different research traditions with their own particular interest such as philosophy of science, laboratory studies, feminist studies, and contributions to risk research (for broad overviews of the field see, e.g., Hackett et al. 2007 and Hess 1997a).

Risk has been, implicitly or explicitly, a recurring theme in STS research. STS is especially known for its critique on the assumption of superiority of science-based knowledge. The underlying assumption of the objective-perceived dichotomy is that “lay persons”

misunderstand the “real” risks as known to science, and thus nonscientific definitions of risks or problems are labeled as “perception” (Wynne 2002b). This implies that risk controversies are envisioned as disagreements between “objective” risk assessments and public misperceptions constructed by ill-informed and emotive publics (Irwin and Wynne 1996). STS scholars criticize this assumed hierarchized dichotomy between experts and laypersons.

STS start from a new understanding of knowledge: they argue that science is a human product, something that has to be “made” or “constructed.” The constructed nature of scientific knowledge is defined in contrast with a naive view of scientific work as a purely rational process of representing nature by using transparent observations (Hess 2001). What becomes “science” or “risk” is the outcome of complex and strategic interactions between divergent actors (Collins 1985). Especially studies of scientific controversies have revealed the complex processes, involving many more actors than just expert scientists, by which reliable knowledge is created and contested (Jasanoff 1999; Shapin and Schaffer 1985).

By emphasizing the constructed nature of science, STS research claims to be impartial with respect to truth or falsity, rationality or irrationality, success or failure of knowledge. STS explain all belief systems symmetrically: they give equal weight to the views of laypeople and experts (Bloor 1976). STS research does not take the views of experts for granted and challenges the assumption that scientific knowledge is the only valid way to discuss risk issues (Wynne 1982).

## Political Sciences

Political science contributes to risk research through its focus on macro-level decision making processes and public policies in the regulation of risks, often comparing different legislations. In policy sciences, the notion “governance” has become a popular concept, referring to a blurring of the state and civil society; to increasing levels of participation; and a shared responsibility between state, business, and civil society (Walls et al. 2005). Policy sciences have an important influence on the development of the idea of risk governance, as governance is a notion directly borrowed from policy science (see section  [Governance](#) for an elaborate discussion of “governance”). Political scientists reflect on changing patterns of governance and differences in regulatory traditions in managing health-related, environmental, and other risks. The focus of their research can vary from questioning how risks are managed in countries with different political environments (e.g., Huitema 2002; Versluis 2003) to comparing the use of the precautionary principle in different regulatory regimes (e.g., Wiener et al. 2010) and analyzing organizational cultures, structures, functions, and processes in controlling and managing risks (e.g., Sparrow 2008).

## Law

The role of law in dealing with the complex dynamics of understanding risks and uncertainty has been one of laying down rules and procedures, for example, on products, substances, or the environment, among which are also the use of scientific advice, participation, and the precautionary principle. Legal research has particularly contributed to a better understanding of how the various institutions ranging from political actors to court deal (and struggle) with risk and uncertainty in specific policy areas with specific reference to the role of the precautionary principle (Alemanno 2007; Arcuri 2007; de Sadeleer 1999; Everson and Vos 2009; Fisher 2008;

Prevost 2008; van Asselt and Vos 2006; Weimer 2010). Within law, the understanding of uncertain risks relates to looking into the regulatory reality (i.e., issues of legality, legitimacy, credibility, and procedural requirements) when decisions must be taken in the face of scientific uncertainty. Most research on uncertain risks in the context of risk governance has focused on genetic engineering, climate change, food technology and safety standards (Rothstein 2009; van Asselt and Vos 2008; Vos 2009; Surdei and Zurek 2009), and environmental risks with much attention for the precautionary principle (Fisher 2008; Vos 1999; Zander 2010).

## Reflection

From this overview, it becomes clear that there are different disciplinary and interdisciplinary approaches and conceptual frameworks to the concept of risk. But together, the social sciences (including contributions from authors originally trained in natural sciences and engineering) have deeply changed our understanding of what “risk” means, “from something real and physical if hard to measure, and accessible only to experts, to something constructed out of history and experiences by experts and laypeople alike” (Jasanoff 1999, p. 150). The different approaches have mostly been positioned with regard to their epistemological premises: a positivistic/realist or a social constructivist view of risk (reviews of the implications of a constructivist versus a realist concept of risk can be found in Bradbury 1989; Jasanoff 1999; Krimsky and Golding 1992; Renn 1992). A realist perspective implies that there is a standard of “real” risk against which lay perception can be measured and shown to be attributed to a lack of public understanding of science and technology. A social constructivist view of risk argues that risk and technology are social processes rather than physical entities, risks do not “simply” reflect the natural reality but are shaped by history, politics, and culture. Public perceptions are therefore not irrational but are as legitimate as other more technocratic views. Even though this categorization makes perfect sense, some approaches are difficult to position and fields can also change their perspective (e.g., SARF is difficult to classify since many scholars started out from a psychometric “realist” perspective but also acknowledge the diversity of risk judgments, see Renn 2008).

Another way to understand the differences in assumption of the diverse fields is to look at whether, and if so, how and which boundaries are assumed, imposed, or contested. The set of boundaries that play a role in risk research are boundaries between subjective and objective risk; between science and nonscience; between experts and laypeople as well as between experts and policymakers; between individual and groups; and between risk assessment and risk management. Drawing or contesting particular boundaries in risk research has a pivotal role in determining the policy problem of risk and a possible framework for solutions (Bradbury 1989). If it is assumed that scientific knowledge is superior to lay persons’ views, then the ensuing solution is a better education of that public. If such boundaries are contested, solutions will be less straightforward and might, for example, call attention to the critical role of experts in political processes.

Where does the “risk governance” framework set out in this chapter stand in this major debate? The framework both “tries to avoid the naïve realism of risk as a purely objective category, as well as the relativistic perspective of making all risk judgments subjective reflections of power and interests.” (Renn 2008, p. 5). Risk governance is about dealing with both the “physical” and “social” dimensions of risk. It expands beyond the dominating technical criteria

for risk analysis and acknowledges public values and concerns as legitimate in their own right but at the same time it searches for ways to benefit from knowledge qualified as scientific. Public views should therefore not be downplayed by labeling them as mere irrational fears. Risk governance is about diffusing boundaries. In this context, we propose to use the notion of “risk perspective” instead of “risk perception” when talking about different viewpoints on risk issues. Risk perception has the long-standing connotation that it implies a distinction between “perceived” (by the “emotive and irrational” public) and “real” (by “objective” scientists) risks (Marris et al. 2001). Thus, a boundary is drawn between one superior knowledge base above another inferior one. “Risk perspective,” in contrast, acknowledges the multiplicity of views on risk in various arenas and in various cultures in a more symmetrical manner (Hermans et al. [in preparation](#)).

## Governance

---

In the previous part, we have discussed ideas that have paved the way for risk governance. Even though many of them do not use the term risk governance explicitly, their contributions have been indispensable in terms of bringing forward ideas, principles, and frameworks for how to deal responsibly with modern risks. The notion of “risk governance” itself has been coined only recently (our discussion of its history follows van Asselt 2007, van Asselt and Renn 2011, and Renn et al. 2011). Risk governance as an emerging concept should be understood in the context of the broader “governance” turn in the policy sciences (Versluis 2003). The notion “governance” came into fashion in the 1980s in circles engaged with development (Stern 2000) and was soon adopted in other domains. The conceptual use of governance has increasingly been adopted in the political science context to emphasize that the state is not the only, single most important actor (there is also a perspective on governance, provocatively termed “governance without government,” see Rosenaau 1995; Rosenaau and Czempiel 1992), which emphasizes the decreased and decreasing role of the nation state) in managing and organizing society. Many classical policy theories share a hierarchic orientation with government as the central actor. In contrast, policy theories inspired by economics award that central role to the market. Both clusters of theories are single-actor in their perspective on power and control. The governance perspective, however, holds that collective binding decisions are generated and implemented in complex multi-actor networks and processes; it also considers various social actors next to state and market such as NGOs and ad hoc coalitions of civilians, of which it is unclear who their supporters are and whom they represent – civil servants, experts, think tanks, agencies, and all kinds of committees active. Power, knowledge, and the capacity to act are distributed among these various actors.

The governance perspective steers away from two other prominent strands of theories, e.g., supranationalism and intergovernmentalism, by raising new questions considering the role and power of states and by drawing attention to the diversity of actors, the diversity of their roles, the manifold relationships between them, and all kinds of dynamic networks emerging from these relationships. When referring to a *multilevel* governance perspective also “government” is no longer a single entity (Rauschmayer et al. 2009). Scholars subscribing to the governance perspective examine actor-networks, the dynamics and the roles of the various actors in these dynamics as a way to understand policy development and political decisions. The shift to governance is best understood as a response to new challenges, such as

globalization, increased international cooperation (such as the European Union), societal changes, including increased citizens engagement, and the rise of nongovernmental organizations (NGOs), the changing role of the private sector, and the augmenting complexity of policy issues. The culmination of all these challenges leads to the need for a new legitimate form of governance (Pierre and Peters 2000; Walls et al. 2005).

The notion governance is used both in a descriptive and in a normative sense (van Asselt 2007; van Asselt and Renn 2011). In a descriptive use of the term, the idea of a complex web of manifold interactions between heterogeneous actors is used to describe the current state of affairs. Governance is then an observation and an approach. The description of governance as “structures and processes for collective decision-making involving governmental and nongovernmental actors” is an example of a descriptive definition (Nye and Donahue 2000). In a normative use, the notion of governance refers to a model or framework for organizing and managing society. In the famous 2001 White Paper of the European Commission on governance (European Commission 2001), such a normative perspective is propagated. In the White Paper, which can be read as a response to the BSE-crisis, governance is presented as an alternative model, in which transparency, stakeholder participation, accountability, and policy coherence are key principles. Often this distinction between description (of the state of affairs) and (policy) model is not made. As a consequence, it is unclear whether governance serves as reference to the framework guiding the analysis or whether it has the status of a (proposed) policy theory.

This is also true for risk governance. Here, the term “governance” is also used in a descriptive and a normative sense. Van Asselt and Renn (2011) argue that on the one hand the state of affairs pertaining to the regulation of many risks is adequately described in terms of governance. Risk decisions can only be understood as the upshot of complex interplays between multiple actors. The governance perspective is needed to sensibly examine and explain the societal dynamics around issues framed as risk issues.

Van Asselt and Renn (2011), furthermore, argue that risk governance also involves the idea that in regulatory practice this state of affairs is not adequately accommodated. The nature of many risks requires cooperation, coordination, and trust between a range of stakeholders, who have diverging interests and different perceptions of the (potential) risks involved. Many risk scholars assert that in case risks are inadequately addressed and managed, this may lead to what the sociologist Ulrich Beck has called “organised irresponsibility” (Beck 1992). Against this background, ideas and principles for how to deal with risks in a more adequate and more responsible manner are proposed. In doing so, governance is no longer used only in a descriptive but also in a normative sense: a new form, or at least new principles, of dealing with risks. Thus, risk governance is a hybrid of an analytical frame and a normative model. Such hybrids are also found in decision theory where the various stages of decision making that the theory suggests can be used as a checklist of how decisions are made (descriptive use) and at the same time functions as a guideline of how to organize the decision process when complex decisions have to be made (normative model) (Keeney 1992, 2004; North 1968).

## **The Origins of Risk Governance**

---

For a detailed discussion see van Asselt (2007) and van Asselt and Renn (2011). The notion “risk governance” has been coined only recently. The origins of the composition of “risk governance” and its introduction to the scholarly literature can be traced back to different

sources. Notable is the link to “TRUSTNET” (European Commission 2000; Heriard Dubreuil et al. 2002) concerted action on “risk governance” (Amendola 2001; Elliot 2001; Heriard Dubreuil et al. 2002), as well as endeavors preceding TRUSTNET. The OECD work on systemic risk (OECD 2003) and the Hood et al. (Hood et al. 2001) book on the government of risk are examples of key trailblazers. In 2001, the first articles with risk governance in their title appeared in two peer-reviewed scientific journals: *Journal of Hazardous Materials* (Heriard Dubreuil 2001) and *Science and Culture* (Elliot 2001). The notion was furthermore used in EU calls for proposals (van Asselt 2007; van Asselt and Renn 2011).

To complicate the matter, interpretations of risk governance differ. In the scholarly literature, risk governance is used as an opposition to the classical notions of risk assessment and risk management by putting uncertainty in center stage and by advancing multi-actor, multifaceted risk processes including contextual factors which together determine the roles, relationships, and responsibilities of particular actors and mechanisms (Renn and Walker 2008; van Asselt and Renn 2011). The European Commission, however, used risk governance more traditionally as an umbrella notion “embracing risk identification, assessment, management and communication” rather than as an alternative paradigm (as cited in van Asselt and Renn 2011). For some years, there was no serious attention given to how risk governance was used and what it actually meant. This situation changed with the foundation of the International Risk Governance Council (IRGC) in 2003 (See ▶ [Box 44.1](#)).

#### Box 44.1. The International Risk Governance Council (IRGC)

The International Risk Governance Council (IRGC), a private, independent, not-for-profit Foundation based in Geneva, Switzerland, was founded in 2003. In the late 1990s, challenges of global technological change such as genetic engineering resulted in increased public concern about risk assessment and management strategies in the EU. During the annual gathering “10th Forum Engelberg” in Switzerland, scientists, government leaders, and heads of industry decided to create an independent, international body to bridge the increasing gaps between science, technological development, decision-makers, and the public.

Since its formal foundation in 2003, it has organized many expert workshops on risk-related issues ranging from critical infrastructures, natural hazards, to nanotechnology and other emerging risks. The IRGC’s work took off with the White Paper No.1 “Risk Governance – Towards an integrative approach” (Renn 2005). This white paper, written under chairmanship of Ortwin Renn, is the first scholarly effort to develop risk governance conceptually. This paper aims to create “an integrated analytic framework for risk governance which provides guidance for the development of comprehensive assessment and management strategies to cope with risks, in particular at the global level. The framework integrates “scientific, economic, social and cultural aspects and includes the effective engagement of stakeholders” (IRGC (International Risk Governance Council) 2005, p. 11). The framework offered two innovations to risk research by including the societal context in the assessment and management of risk, and by categorizing risks based on the knowledge about it, distinguishing between “simple,” “complex,” “uncertain,” and “ambiguous” risks (▶ [Fig. 44.1](#)).

This framework displays the five key elements which reflect the way in which risk can be dealt with that fully accounts for the societal context of both the risk and the decision that is reached (IRGC 2005).

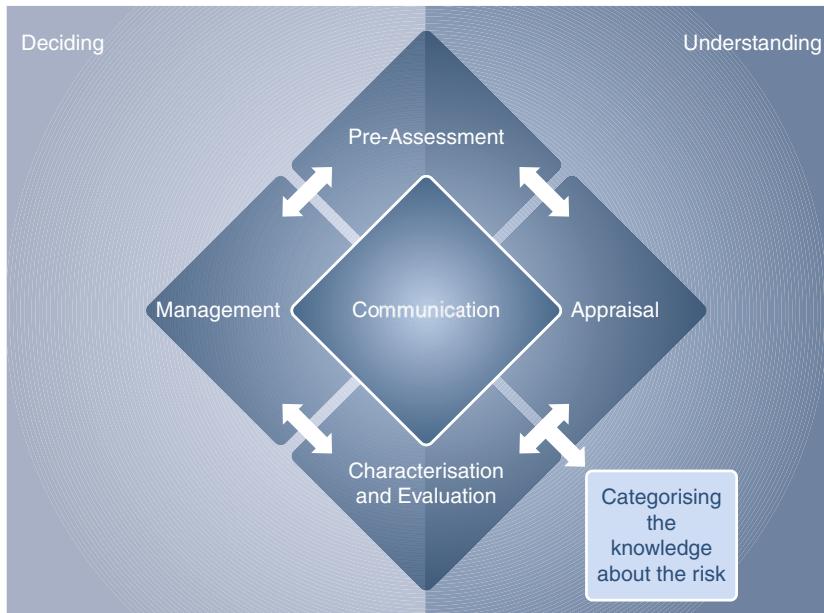


Fig. 44.1

The IRGC risk governance framework. (Source: <http://www.irgc.org/-The-IRGC-risk-governance-framework,82-.html>, with kind permission from the International Risk Governance Council)

Through a network between academia, NGOs, regulators, and industry, it is the IRGC's aim to jointly achieve coordinated and coherent policy making, regulation, research agendas, and communication with regard to the governance of risks.

The White Paper was also published as a lead chapter in the books *The Tolerability of Risk management* (Bouder et al. 2007) and *Global Risk Governance: Concept and Practice Using the IRGC Framework* (Renn and Walker 2008); see also a similar chapter in Bischof (2008), in which next to the framework itself critical reviews and case studies have been included. Subsequently, some agencies and national regulatory bodies have partially adopted the framework and designed manuals on how to use it for their specific purpose (see, e.g., Dreyer and Renn 2009 for EFSA and the Handbook for Risk Assessment and Policy Advice of the Dutch Food and Consumer Product Safety Authority (Voedsel en Waren Autoriteit (VWA) 2010).

## The Typology of Risk

For the typology of risk, we adopt the approach of van Asselt and Renn (2011) and Renn et al. (2011). This typology is a further development of Renn's original typology as set out in Klinke and Renn (2002) and the IRGC White Paper on Risk Governance (2005). Central to risk governance is the recognition that there are various types of risks. Since the Knightian definition (see section **Risk** of this chapter), risks have been treated in terms of probability and effects, dose and response, and agent and consequences. Risk governance involves the

recognition that uncertainty and risk cannot as easily be distinguished as is assumed in the positivistic risk paradigm. Some risks are simple, namely, calculable and relatively easy to manage. In those cases, past experience and the associated availability of statistical data enable to estimate probability and to derive a measure of effect. Existing risk assessment tools and risk management approaches suffice. Examples involve car accidents and regularly recurring natural events, such as seasonal flooding.

However, many risks cannot be classified as “simple.” They are not confined to national borders or a single sector, and do not fit the linear, mono-causal model of risk. Instead, the analysis must focus on interdependencies and ripple and spill-over effects that initiate impact cascades between otherwise unrelated risk clusters (Hellstroem 2001; van Asselt and Renn 2011). A well-known example is BSE which had effects not only on the farming industry but also on the industry of animal feed, the economy as a whole, and politics (Vos 2000; de Bandt and Hartmann 2000; OECD 2003; Renn and Keil 2009; see ➤ **Box 44.2** in this chapter). The transmission effects were globally diffused to all areas of the world, including those that were

#### **Box 44.2. The Impact of the BSE Crisis in Risk Research**

The BSE crisis has provided a turning point in the way actors – industry, regulators, scientists, and many others – have started to deal with risks and uncertainties. Prior to the BSE crisis, the EU managed risks in, e.g., food safety regulation on a rather ad hoc basis (Vos 2000). The BSE case demonstrated the need for a more structural regulation in policy fields in which risks play a role, such as food safety, environment health and safety, and chemical policy. BSE was first identified in Britain in the mid-1980s, and is thought to be caused by the remnants of slaughtered animals that have ended up in the high-protein diets of cattle in the beginning of the 1980s. By eating this, cows could develop a fatal neurodegenerative disease, commonly known as “Mad Cow Disease.” Although it has not been scientifically verified, it is thought that by either eating infected beef or by inhaling the bone meal fertilizer, people risk infection and could develop a variant of the disease, Creutzfeld Jacob, which is a fatal neurological disorder.

The initial response to this uncertain risk by regulators – seeking certainty about the parameters of the crisis – illustrated the need for a new way of approaching risk-related issues. Although research was ongoing, regulators attempted to obtain plausibility proofs, i.e., they increasingly resorted to science for more certainty and conclusive evidence. They hoped that this would sooth an anxious public and would allow for a quick fix to a persistent problem full of uncertainty that was providing a threat to the economy since beef sales plummeted. In 1990, in an ultimate attempt to reassure the public, restore trade in beef, and establish regulatory credibility, the British Minister of Agriculture, well aware of the uncertainty of the risk, fed his 4-year-old daughter a hamburger on British national television, implying its absolute safety. Several years thereafter, scientists discovered a possible link between mad cow disease and the human Creutzfeld Jacob variant.

The aftermath of the BSE crisis led to serious rethinking of risk regulation in academia and beyond and was an incentive for many to develop and advocate a paradigm shift toward a more inclusive way of dealing with risks. It became a “textbook example” to show that the way in which certain risks – such as BSE – were assessed and managed were no longer adequate or acceptable. Moreover, it also led to reforms and/or the birth of major risk regulating institutions in several European countries (e.g., the UK) and the EU (with the formation of a new “European Food Safety Authority,” EFSA) (Oosterveer 2002; Renn 2008; van Zwanenberg and Millstone 2005; Vos 2000).

not immediately affected by the crisis. Such risks are complex (multi-causal) and surrounded by uncertainty and/or ambiguity (Renn et al. 2011; van Asselt and Renn 2011). It is difficult to identify, let alone quantify, multi-causal, usually nonlinear, links between a multitude of potential causal agents and specific effects. Complexity can be caused by interactive effects among agents (synergisms or antagonisms), long delay periods and the associated latency lacunae (this notion has been introduced to the risk literature by Harremoës et al. 2001 – latency lacuna refers to the fact that technologies are improved during the period in which health and/or environmental impacts are studied; when such monitoring and impact studies identify risks, the question is whether those findings still hold for the newer generation of the technology), interindividual variation, etc. Due to complexity, it is impossible to achieve complete deterministic knowledge of cause–effect relationships.

Risk refers to the possibility of damage, whether in health, environmental, economic, or other terms. As long as the risk has not manifested itself in damages, the threat is potential and is evaluated by one or more actors as negative. Due to the fact that these situations often involve structural changes and or new hazards, they are highly uncertain. Uncertainty is not simply the absence of knowledge (compare van Asselt 2000 and Levidow et al. 2005). However, addressing uncertainty is a challenging, far from straightforward, job. Numerous scholars agree that there cannot be a single approach in addressing uncertainty that will satisfy in all circumstances and contexts (Bailey et al. 1996; Funtowicz and Ravetz 1990; Harremoës, et al. 2001; Health Council of the Netherlands 2008; Klinke and Renn 2002; Morgan and Henrion 1990; O’Riordan and McMichael 2002; Pollack 2003; Ravetz (1996 [1971]); van Asselt and Petersen 2003; van Asselt and Renn 2011; van Asselt and Rotmans 2002; van der Sluijs 1997; Walker et al. 2003; WRR (Scientific Council for Government Policy) 2010).

In addition to complexity and uncertainty, risk governance includes a third component: ambiguity (Klinke and Renn 2002; Renn and Roco 2006; van Asselt and Renn 2011). Ambiguity refers to the existence of multiple values. Ambiguity results from divergent and contested perspectives on the justification, severity, or wider meanings associated with a perceived threat (compare Stirling 2003). As a consequence, views differ on the ways to assess and appraise the risks, and more in particular on the relevance, meaning, and implications of available risk information and on which management actions should be considered. This means that there are different legitimate viewpoints from which to evaluate whether there are or could be adverse effects and whether these risks are tolerable or even acceptable. Risks are acceptable in case they are considered low or nonexisting, so additional regulatory efforts are considered unnecessary. Activities are tolerable if they are considered as worth pursuing for the benefit that they carry (Bouder et al. 2007). In cases of tolerable risks, additional regulatory efforts for risk reduction or coping are welcomed. Actors, however, respond to risks according to their own risk constructs and images, yielding several meaningful and legitimate interpretations of risk assessment outcomes (Keeney 2004). As a consequence, whether risks are acceptable, tolerable, or not can be subject of considerable debate and intense controversy. Ambiguity is used to refer to such social situations around risk issues. Examples involve controversies pertaining to passive smoking (although the health risks of active smoking are uncontroversial).

Many of the non-simple risks discussed pertain to future risks, namely, potential hazards that may or may not result in damage in the long run. Take the example of nanotechnology: the risk assessment for this new technological development depends on theoretical, nonempirical insights and ideas about causal relationships between exposure(s) and effect(s) on human health and environment. Furthermore, it depends on the decisions that humans make about

the use, application, exposure barriers, and safety culture with respect to these technologies. Finally, the context, such as the level of trust in the regulators, major accident(s) elsewhere, and coincidences between exposure to nanoparticles and detrimental effects is of influence.

Another illustration is the case of wireless communication technology, a popular and ubiquitous technology with a very high penetration rate that nevertheless ignites public concern, especially at a local level where the technology is implemented with thousands of base stations (see, e.g., Burgess 2004; Soneryd 2007; Stilgoe 2007). Governments attempt to act responsibly, but are confronted with little, inherently uncertain evidence that this technology poses a threat to human health. While the majority of experts emphasize that to date no consistent evidence has demonstrated cancer risks, uncertainties remain about long-term effects and effects on children as well as other health effects (van Asselt et al. 2009). A growing literature has illustrated the difficulty in dealing with such risks, since the nature of the risk is the outcome of a complex interplay of science, technology, and society (e.g., Jasanoff 2005).

Van Asselt and Renn (2011) and Renn et al. (2011) argue that it is possible, in theory, to distinguish between uncertain, complex, and ambiguous risks. However, uncertainty often results from complexity (van Asselt 2000). An illustrative example is the case of the introduction of genetically modified species in the environment. It is generally accepted that the risks to the environment and/or human beings are highly uncertain. Krayer von Kraus (2005) analyzed how experts view such risks. From his analysis detailing the varying ideas on which variables matter and which mechanisms should be included in the causal scheme, it is also clear that GMOs constitute an example of complex risks. Furthermore, taking into account that the risks are evaluated differently by different experts, as has been convincingly demonstrated by Jasanoff (2005), it is clearly also an example of ambiguous risks. So risks associated with genetic modification, and agro-biotechnology in particular, are best characterized as risks that are uncertain, complex, and ambiguous. The same is true for such risks such as nuclear energy or climate change.

Uncertainty, complexity, and ambiguity point to different reasons why many risks defy simple concepts of causation (van Asselt and Renn 2011). Each of the three characteristics of risks contributes to a better understanding of the situation in which risks emerge and manifest themselves. Risk governance thus highlights the importance of uncertain, complex, and/or ambiguous risks. In its 2008 scientific report “Uncertain Safety,” the WRR (the Dutch Scientific Council for Government) calls for a paradigm shift with regard to the governance approach to risks. The WRR (WRR (Scientific Council for Government Policy) 2010) as well as the Health Council in the Netherlands (Health Council of the Netherlands 2006, 2008), for instance, have made an effort to translate the ideas that pertain to risk governance, among which is the acceptance of uncertainty, to practice. These scientific advisory bodies play a key role in advising the Dutch government about societal relevant issues. They furthermore aim to form an intersection in international policy (<http://www.wrr.nl/english/>, accessed April 5, 2011). They argue that the classical risk paradigm and its policy based on “simple” risks are outdated, but should not disappear. Rather, a paradigm shift to risk governance should take place, focusing policy on uncertain, complex, and ambiguous risks. Simple risks, which inhibit little to no uncertainty (WRR (Scientific Council for Government Policy) 2010), the so-called certain uncertainties (van Asselt 2000), should have the status of the special cases, rather than the dominant position that they have in current policy on risk management and assessment practices.

Furthermore, it is a consistent finding in the body of literature discussed in section  **The Origins of Risk Governance** that very often, uncertain, complex, and/or ambiguous risks are

treated, assessed, and managed as if they were simple (Renn et al. 2011; van Asselt and Renn 2011). The assessment and management routines in place do not do justice to the nature of such risks. The consequences of this maltreatment range from social amplification or irresponsible attenuation of the risk to sustained controversy, deadlocks, legitimacy problems, unintelligible decision-making, trade conflicts, border conflicts, expensive rebound measures, and lock-ins. The main message from risk governance is that it is urgently needed to develop better conceptual and operational approaches to understand and characterize non-simple risks.

## Risk Governance

---

Risk governance or the *new risk approach*, as it is called by the WRR, has gradually been created, developed, and discussed in scientific literature and has slowly entered the organizational level as well (WRR (Scientific Council for Government Policy) 2010). However, regulatory and managerial understanding, let alone practical application and implementation of this approach, still need development. Complex, uncertain, and ambiguous risks require an organizational setting which fosters an interdisciplinary perspective, flexibility, and diversity, which is at odds with current managerial practices (WRR (Scientific Council for Government Policy) 2010; Health Council of the Netherlands 2008; van Asselt and Renn 2011).

What is needed to treat uncertain, complex, and/or ambiguous risks adequately? Van Asselt and Renn (2011) argue that first of all, it is important to accept scientific uncertainty and controversy and public debate as the state of affairs. In many cases the governing of risks will involve precaution in the sense of a cautious and flexible strategy that enables learning from restricted errors, new knowledge and visible effects, so that adaption, reversal, or adjustment of regulatory measures is possible (See also De Vries et al. 2011; van Dijk et al. 2011). Precaution also entails the responsibility for early warning and monitoring in order to facilitate systematic searching for new hazards by institutions of government, business, or civil society (Charnley and Elliott 2002). Risk governance is not just concerned with minimizing risks, but also with stimulating resilience (or decreasing vulnerability) in order to be able to withstand or even tolerate surprises (Collingridge 1996).

## Risk Governance Principles

---

Risk governance endorses highly contextualized practices of dealing with risks; it is not a model in the strict sense of the word. The idea of risk governance aims to serve a paradigm shift that helps risk professionals to familiarize themselves with a broader concept of risk. Van Asselt and Renn (2011) and Renn et al. (2011) proposed to synthesize the various ideas and proposals in a set of principles, which can inform about how to deal with uncertain, complex, and/or ambiguous risks in various contexts:

- Communication and inclusion
- Integration
- Reflection

The set of principles are discussed in more detail below.

## The Communication and Inclusion Principle

In the context of risk governance, van Asselt and Renn (2011) argue that communication is crucial. Effective mutual communication takes center stage in the challenges risk governance aims to address and should be approached accordingly. In case of sidestepping communication, successful risk governance is irretrievably harmed. As we discussed in section [The Origins of Risk Governance](#), initially, risk communication has been approached in terms of educating and persuading the public (Fischhoff 1995). However, this “deficit model” has been questioned by research on risk controversies (see, e.g., Horlick-Jones 1998; Irwin and Wynne 1996) that shows that the public is often falsely dismissed as a collection of laypersons incapable of understanding and interpreting science. Risk governance builds on the acknowledgment that there are various, conflicting risk perspectives.

Van Asselt and Renn (2011) refer to communication as meaningful interactions in which knowledge, experiences, interpretations, concerns, and perspectives are exchanged (compare Löfstedt 2005). The role of communication within risk governance is threefold. To begin with, communication entails the process of sharing knowledge and information on the various risk perspectives. Secondly, it may lead to the inclusion of various actors in the decision-making process which will lead to a sense of ownership. Communication might under certain conditions simultaneously increase the level of trust among all actors involved (Löfstedt 2005) which is an important, if not necessary, component in the acceptance of particular risk management arrangements.

However, communication in the context of risk governance is not simple. It is not just a matter of bringing people together. Social learning is required in order to find ways to discuss risk perspectives. Preparing the structure of the process is key. It is also required to figure out which type of communication with whom is important in which phase. Constructive communication does not imply that all actors remain in a constant dialogue with each other. It also does not imply that this question of “who is important in which phase” can be easily answered. Rather, communication requirements may differ depending on the context, such as political culture, the dominant social values and the trust relationships between actors. Hence, enabling communication is insufficient. The interaction level of the actors involved needs to tie into the challenges that accompany uncertainty, complexity, and ambiguity. By keeping a constant eye on “who is important in which phase,” it will become visible throughout the process which type of communication with whom is constructive and contributes to the responsible governance of uncertain risks. This remains a trial and error process in which various actors learn.

Inclusion has deep implications. Contrary to the current state of affairs in which risk topics are usually identified by experts, risk perspectives of other actors may act as the driving agents for identifying risk topics. Inclusion does not just mean that various actors are included, but that they play a key role in framing (or pre-assessing) the risk (IRGC (International Risk Governance Council) 2005; Renn 2008; see also Roca et al. 2008). Inclusion should be open and adaptive at the same time (Stirling 2004). Inclusion is defended for several reasons (compare Renn et al. 2011; Roca et al. 2008; van Asselt and Renn 2011). First, inclusion is needed to explore various sources of information and knowledge and to identify various risk perspectives. Second, it is argued from a democratic perspective that actors affected by the risks and/or the ways in which the risks are governed have a right to participate in deciding about those risks. Thus, inclusion is not just a means, but an end in itself. Third, it is argued that the more actors are involved in weighing the essentially heterogeneous pros and cons, the more socially robust

the outcome. People engaged in the participatory process tend to be more satisfied about the process itself than about its outcomes. Inclusion is thus supposed to support the coproduction of risk knowledge, the coordination of risk evaluation, and the design of risk management.

However, including various actors can be a challenge. The challenge is to organize productive and meaningful communication with, and inclusion of, a range of actors, who have complementary roles and diverging interests. The available empirical analyses suggest that the attempt to include different stakeholders can help to de-escalate conflicts and to legitimize the final decision that will always disappoint some actors in society (Beierle and Cayford 2002; US-National Research Council of the National Academies 2008). Inclusion does not, however, necessarily reduce conflict or lead to more widely accepted decisions (Kinney and Leschine 2002). Participation procedures themselves can become a source of conflict (Wiedemann and Femers 1993). Not every relevant actor might be interested in participating. Some actors might try to impose their framing on the process from the very beginning (van Asselt and Vos 2006). It is important to accept and address conflict, even though in many cases conflicts cannot be settled nor should that be the aim.

## The Integration Principle

Integration refers to the need to collect and synthesize all relevant knowledge and experience from various disciplines and various sources, including uncertainty information and articulations of risk perspectives. Scientific expertise should therefore not be regarded as a panacea to provide clear-cut solutions to non-simple risk problems. Risk governance still seeks scientific knowledge, but in order to look beyond the terms of likelihood and effects when it concerns uncertain, complex, and/or ambiguous risks. The integration principle emphasizes that also values and issues such as reversibility, persistence, ubiquity, tolerability, equity, catastrophic potential, controllability, and voluntariness should be integrated in risk assessment and evaluation. Furthermore, risk governance is not just about risks and usually not about a single risk. Risk governance requires risk(s)–benefit(s) evaluations and risk–risk trade-offs. The integration principle reflects the importance of such multidimensional evaluations.

Although it is quite impossible to ever fully understand uncertain, complex, and/or ambiguous risks, improvements can be made to understand them better. One way of doing this is, on the one hand, by transcending disciplinary (academic) boundaries, and on the other hand, by including more practical and tacit knowledge, in order to reflect a large variety in social and cultural values, preferences, and worldviews. Such an extended perspective will enable a set of consistent and coherent scenarios of future decision opportunities on which the relevant actors can make informed choices (see for an example of such regulation scenarios Fox et al. 2011).

Integration also refers to the process itself. Risk governance advances a holistic approach to framing, appraising, characterizing, evaluating, and managing risks (Zinn and Taylor-Gooby 2006). This implies that a strict separation between risk assessment and risk management is counterproductive and in need of critical reexamination (Jasanoff 1993). Risk governance is not a linear, sequential three-stage process of risk assessment, management, and communication, but it is dynamic and it requires interlinked and iterative processes. Although it may still be useful to distinguish assessment (examining the risks and benefits) from management (identifying regulatory options), it is important to realize that they can and should not be viewed as unconnected activities to be carried out in different realms. In other words, the separation of

risk assessment and risk management does not imply a complete “divorce” (Williams and Thompson 2004, p. 1622). Jasianoff (2005) as well as van Asselt and Vos (2008) have argued that boundary work is an effective way to make problems appear to be simple. Defining risk assessment and management as separate realms enables analysts to ignore uncertainty, complexity, and ambiguity. The integration principle calls attention to the need to consider the interconnections, both content-wise and in terms of process, between the various risk-related activities. Shedding light on these interconnections may also enable the actors involved to gain a better overview of the risks and uncertainties involved, which may lead to better practices.

## The Reflection Principle

Unfortunately, risk governance cannot be routinized. Differentiation is not an exception, but rather the rule. Reflection is a necessary component since it enables actors and institutions to maintain a critical outlook on what they are doing (compare Beck et al. 1994; Schon 1983) in order to continue to treat the risks as uncertain, complex, and/or ambiguous instead of simple, which each require different practices (van Asselt and Vos 2006, 2008; Wynne 2002a). What is needed is a collective reflection about balancing the possibilities for overprotection and underprotection. Van Dijk et al. (2011) refer to this balancing act as “prudent precaution.” If too much protection is sought, innovations may be prevented or stalled; if too little protection is provided, society may experience unmanageable unpleasant surprises. The classic question “How safe is safe enough?” is replaced by the question “How much uncertainty is the collective willing to accept in exchange for some benefit(s)?” So the focus shifts from risk to uncertainty (compare De Vries et al. 2011). The communication and inclusion principle hold that various actors take part in this reflective discourse and discuss how decisions could and should be made in the face of irresolvable uncertainty, complexity, and ambiguity. The reflection principle emphasizes that there are important difficult issues (uncertainty, complexity, ambiguity) which are in need of repeated consideration of all actors throughout the process. Every situation requires differentiation and flexibility and becomes a balancing act. Otherwise, the process risks to (re)introduce the familiar frames and routines developed for simple risks. Hence we have to be alert.

A crucial component of such reflexivity is to remain critical about inclusion and integration. One should not aim for a situation in which there is full trust – even though this might occasionally be achieved – since it often implies an approach aimed at risk acceptance rather than an approach aimed at critical reflection from public skepticism. Moreover it is important to find and maintain a pragmatic balance, whereby each risk is assessed on its own characteristics within a certain context, thereby taking into account its associated social dynamics (Walls et al. 2005). Risk governance thus may contribute to a “repolitization” of risk questions that have been “depolitized,” i.e., risk issues that have only been treated technocratically. In other words, depolitization is not the strategy behind risk governance (compare van Asselt and Vos 2006 who explicitly call attention to the political deficit in current ways of dealing with uncertain, complex, and/or ambiguous risks). Risk governance thus fundamentally differs from the traditional, positivistic approach that aims at depolitization by means of technocratization and strict boundaries between realms and activities. Reflexivity requires more scholarly consideration as well, since the contemporary scholarly debate seems to focus mainly on the communication and inclusion principles.

In sum, risk governance should not be considered as a panacea. It rather aims to facilitate a broader understanding of issues pertaining to contemporary risks. Risk governance aims to integrate insights that can be gained from decades of research on risk. The most recent development is to attempt to synthesize the insights gained into a set of principles. This set of principles has to prove to be both robust and applicable to practice.

## Conclusion and Discussion

---

In this chapter, in order to contribute to risk theory, we have explored and analyzed the origins and contours of risk governance. To that end, we reflected on important scientific developments in different strands of literature that have all contributed to the current understanding and thus explicitly or implicitly to the idea of risk governance in risk theory. These contributions are very multidisciplinary, ranging from psychology to law to science and technology studies (STS). We also discussed risk governance as part of the broader governance turn in policy sciences. Following the state-of-the-art review by van Asselt and Renn (2011), we emphasized that it is central to risk governance that it is recognized that uncertainty and risk cannot as easily be distinguished as is assumed in this positivistic risk paradigm. Many risks which require societal choices and decisions can be adequately characterized as complex, uncertain, and/or ambiguous risks. Risk governance pleads that uncertain, complex, and/or ambiguous risks are recognized and it aims to provide a basis for (more) adequate treatment. It is a consistent finding in the risk literature, that too often these risks are treated, assessed, and managed as if they were simple. The persistent societal controversies pertaining to genetic engineering, nuclear energy, and chemical risks suggest an urgent need to develop alternative approaches to deal with uncertain, complex, and/or ambiguous risks. We have discussed risk governance as an attempt to provide a basis for such alternatives. We discussed the proposal to see risk governance as a set of principles that can inform thinking on non-simple risks practices: the communication and inclusion principle, the integration principle, and the reflection principle. These principles aim to synthesize the most important aspects to be organized in order to be able to govern risks responsibly. The multitude of references to “risk governance” has unduly given it the status of a “buzzword.” However, we believe that it should be understood as a plea for a paradigmatic and practical shift in dealing with modern risks. But it will not be an easy passage. We think that taking stock might help to facilitate a shift in risk practices.

## References

---

- Adams J (2005) *Risk*. UCL Press, London
- Alemanno A (2007) Trade in food: regulatory and judicial approaches in the EC and the WTO. CMP Publishing, London
- Amendola A (2001) Recent paradigms for risk informed decision making. *Safety Sci* 40(1–4):17–30
- Arcuri A (2007) Reconstructing precaution: deconstructing misconception. *Ethics Int Aff* 21(3):359–379
- Bailey P, Gough C, Chadwick M, McGranahan G (1996) Methods for integrated environmental assessment: research directions for the European Union. Stockholm Environment Institute, Stockholm, Sweden
- Beck U (1992) *Risk society. Towards a new modernity*. Sage, London
- Beck U (2009) *World at risk*. Polity, Cambridge
- Beck U, Giddens A, Lash S (1994) *Reflexive modernization: politics, tradition and aesthetics in the modern social order*. Polity, Cambridge
- Beierle TC, Cayford J (2002) *Democracy in practice: public participation in environmental decisions*. Resources for the Future Press, Washington, DC

- Bernstein PL (1996) Against the gods. The remarkable story of risk. Wiley, New York City
- Bischof H-J (ed) (2008) Risks in modern society. Springer, Berlin/Heidelberg
- Bloor D (1976) Knowledge and social imagery. Routledge & Kegan Paul, London
- Bouder FE (2008) Defining a tolerability of risk framework for pharmaceutical products in a context of scientific uncertainty. King's College, London
- Bouder FE, Slavin D, Löfstedt R (eds) (2007) The tolerability of risk. A new framework for risk management. Earthscan, London
- Bradbury JA (1989) The policy implications of differing concepts of risks. *Sci Technol Hum Values* 14(4): 380–399
- Burgess A (2004) Cellular phones, public fears, and a culture of precaution. Cambridge University Press, Cambridge
- Charnley G, Elliott ED (2002) Risk versus precaution: environmental and public health protection. *Environ Law Rep* 32(2):10363–10366
- Collingridge D (1996) Resilience, flexibility, and diversity in managing the risks of technologies. In: Hood C, Jones DKC (eds) Accident and design: contemporary debates in risk management. UCL Press, London, p 4045
- Collins HM (1985) Changing order. Replication and induction in scientific practice. Sage, London
- de Bandt O, Hartmann P (2000) Systemic risk: a survey. ECD Working Paper nr. 35, available at <http://ssrn.com/abstract=258430>
- de Sadeleer N (1999) Het voorzorgsbeginsel: Een stille revolutie. *TMR* 8:82–99
- De Vries G, Verhoeven I, Boeckhout M (2011) Taming uncertainty: the WRR approach to risk governance. *J Risk Res* 14(4):485–499
- Douglas M, Wildavsky A (1982) Risk and culture. University of California Press, Berkeley
- Dreyer M, Renn O (eds) (2009) Food safety governance. Integrating science, precaution and public involvement. Springer, Heidelberg/New York
- Elliot D (2001) Risk governance: is consensus a con? *Sci Cult* 10(2):265–271
- European Commission (2000) The TRUSTNET framework: a new perspective on risk governance. [www.trustnetinaction.com](http://www.trustnetinaction.com). Accessed 21 June 2010
- European Commission (2001) European governance: a white paper. Commission of the European Communities, Brussels
- Everson M, Vos E (eds) (2009) Uncertain risks regulated. Routledge-Cavendish, Oxon
- Fischer E (2002) Precaution, precaution everywhere: developing a 'common understanding' of the precautionary principle in the European Union. *Maastricht J European Compar Law* 9(1):7–28
- Fischer E, Jones JS, von Schomberg R (eds) (2006) Implementing the precautionary principle: perspectives and prospects. Edward Elger, Cheltenham/Northampton
- Fischhoff B (1995) Risk perception and communication unplugged: twenty years of process. *Risk Anal* 15(2):137–145
- Fischhoff B, Watson SR, Hope C (1984) Defining risk. *Policy Sci* 17(2):123–139
- Fisher E (2008) Opening pandora's box: contextualising the precautionary principle in the European Union. In: Vos E, Everson M (eds) Uncertain risks regulated. Routledge-Cavendish, Oxon
- Fox T, Versluis E, van Asselt MBA (2011) Regulating the use of bisphenol a in baby and children's products in the European Union: current developments and scenarios for the regulatory future. *European J Risk Regul* 2(1):21–35
- Funtowicz SO, Ravetz JR (1990) Uncertainty and quality in science for policy. Kluwer, Dordrecht
- Funtowicz SO, Ravetz JR (1992) Risk management as a postnormal science (comment). *Risk Anal* 12(1): 95–97
- Giddens A (1991) Modernity and self-identity. Self and society in the late modern age. Polity, Cambridge
- Hackett EJ, Amsterdamska O, Lynch M, Wajcman J (eds) (2007) The handbook of science and technology studies. MIT Press, Cambridge/London
- Harremoës P, Gee D, MacGarvin M, Stirling A, Keys J, Wynne B et al. (2001) Late lessons from early warnings: the precautionary principle 1896–2000 (No. Environmental Issue Report 22). European Environment Agency, Copenhagen
- Health Council of the Netherlands (2006) Health significance of nanotechnologies. Health Council of the Netherlands, The Hague
- Health Council of the Netherlands (2008) Prudent precaution. Health Council of the Netherlands, The Hague
- Hellstroem T (2001) Emerging technological and systemic risk: three cases with management suggestions. Contribution to the OECD international futures project on emerging systemic risks. OECD, Paris
- Heriard Dubreuil GF (2001) Present challenges to risk governance. *J Hazard Mater* 86(1–3):245–248
- Heriard Dubreuil GF, Bengtsson G, Bourrelier PH, Foster R, Gadbois S, Kelly KN et al (2002) A report of TRUSTNET on risk governance: lessons learned. *J Risk Res* 5(1):83–95
- Hermans MA, van Asselt MBA, Passchier WP (in preparation). Conceptualizing risk controversies. A review of social science research on wireless communication technology (working title)
- Hess D (1997a) If you're thinking of living in STS... a guide for the perplexed. In: Downey G, Dumit J,

- Traweek S (eds) *Cyborgs and citadels*. School for American Research Press, Santa Fe
- Hess JD (1997b) *Science studies. An advanced introduction*. New York University Press, New York/London
- Hess D (2001) Ethnography and the development of science and technology studies. In: Atkinson P, Coffey A, Delamont S, Lofland J, Lofland L (eds) *Handbook of ethnography*. Sage, London, pp 234–245
- Hood C, Rothstein H, Baldwin R (2001) *The government of risk: understanding risk regulation regimes*. Oxford University Press, Oxford
- Horlick-Jones T (1998) Meaning and contextualization in risk assessment. *Reliab Eng Syst Saf* 59:79–89
- Huitema D (2002) Hazardous decisions. Hazardous waste siting in the UK, the Netherlands and Canada. Institutions and discourses. Kluwer, Dordrecht
- IRGC (International Risk Governance Council) (2005) *Risk governance: towards an integrative approach*. IRGC, Geneva
- IRGC (International Risk Governance Council) (2007) *An introduction to the IRGC risk governance framework*. IRGC, Geneva
- Irwin A, Wynne B (eds) (1996) *Misunderstanding science. The public reconstruction of science and technology*. Cambridge University Press, Cambridge
- Jasanoff S (1990) *The fifth branch: science advisers as policy makers*. Harvard University Press, Cambridge, MA
- Jasanoff S (1993) Bridging the two cultures of risk analysis. *Risk Anal* 13(2):123–129
- Jasanoff S (1999) The songlines of risk. *Environ Value* 8:135–152
- Jasanoff S (2005) *Designs on nature. Science and democracy in Europe and the United States*. Princeton University Press, Princeton
- Kasperson JX, Kasperson RE (1991) Hidden hazards. In: Mayo DG, Hollander RD (eds) *Acceptable evidence: science and values in risk management*. Oxford University Press, Oxford
- Kasperson JX, Kasperson RE (2005a) *The social contours of risk*, vol I. Earthscan, London/Sterling
- Kasperson JX, Kasperson RE (2005b) *The social contours of risk: volume I: publics, risk communication and the social amplification of risk*. Earthscan, London/Sterling
- Kasperson RE, Renn O, Slovic P, Brown HS, Emel J, Goble R et al (1988) The social amplification of risk: a conceptual framework. *Risk Anal* 8(2):177–187
- Keeney RL (1992) *Value-focused thinking. A path to creative decision making*. Harvard University Press, Cambridge
- Keeney RL (2004) Making better decision makers. *Decis Anal* 1(4):193–204
- Kinney AG, Leschine TM (2002) A procedural evaluation of an analytic-deliberative process: the Columbia river comprehensive impact assessment. *Risk Anal* 22(1):83–100
- Klinke A, Renn O (2002) A new approach to risk evaluation and management: risk-based, precaution-based, and discourse-based strategies. *Risk Anal* 22(6):1071–1094
- Knight FH (1921) *Risk, uncertainty and profit*. Houghton Mifflin, Boston
- Krayer von Kraus MP (2005) Uncertainty in policy relevant sciences. Technical University of Denmark, Copenhagen
- Krimsky S, Golding D (eds) (1992) *Social theories of risk*. Praeger, Westport
- Levidow L, Carr S, Wield D (2005) European Union regulation of agri-biotechnology: precautionary links between science, expertise and policy. *Sci Public Policy* 32(4):261–276
- Löfstedt RE (2005) *Risk management in post-trust societies*. Palgrave Macmillan, Hampshire/New York
- Löfstedt RE, Renn O (1997) The Brent Spar controversy: an example of risk communication gone wrong. *Risk Anal* 17(2):131–136
- Marris C, Wynne B, Simmons P, Weldon S (2001) Public perceptions of agricultural biotechnologies in Europe (No. Final Report of the PABE research project funded by the Commission of European Communities (Contract number: FAIR CT98-3844 (DG12 -SSMI)). UK Lancaster University, Lancaster
- Maynard Keynes J (1921 [2004]) *A treatise on probability*. MacMillan, London
- Morgan GM, Henrion M (1990) *Uncertainty-a guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge University Press, New York
- Mourik R (2004) Did water kill the cows? The distribution and democratisation of risk, responsibility and liability in a Dutch agricultural controversy on water pollution and cattle sickness (PhD). Pallas Publications, Maastricht
- North DW (1968) A tutorial introduction to decision theory. *IEEE T Syst Sci Cyb* 4(3):200–210
- Nowotny H (1976) Social aspects of the nuclear power controversy. IIASA, Laxenburg
- Nowotny H (2008) *Insatiable curiosity: innovation in a fragile future*. MIT Press, Cambridge, MA
- Nowotny H, Scott P, Gibbons M (2001) *Re-thinking science: knowledge and the public in an age of uncertainty*. Polity Press in association with Blackwell Publishers, Cambridge
- Nye JS, Donahue JD (eds) (2000) *Governance in a globalising world*. Brookings Institution, Washington, DC
- O'Riordan T (1982) Risk perception studies and policy priorities. *Risk Anal* 2(2):95–100
- O'Riordan T, McMichael AJ (2002) Dealing with scientific uncertainties. In: Martens P, McMichael AJ (eds)

- Environmental change, climate and health: issues and research methods. Cambridge University Press, Cambridge
- O'Riordan T, Cameron J, Jordan A (2001) Reinterpreting the precautionary principle. *Cameron May International Law & Policy*, London
- OECD (2003) Emerging systemic risks. Final Report to the OECD futures project. OECD, Paris
- Oosterveer P (2002) Reinventing risk politics: reflexive modernity and the European BSE crisis. *J Environ Policy Plann* 4(3):215–229
- Pidgeon NF, Kaspelson RE, Slovic P (eds) (2003) The social amplification of risk. Cambridge University Press, Cambridge
- Pierre J, Peters BG (2000) Governance, politics and the state. Macmillan, Houndsmill/London
- Pollack HN (2003) Uncertain science... uncertain world. Cambridge University Press, Cambridge
- Prevost D (2008) Private sector food-safety standards and the SPS agreement: challenges and possibilities. *South African Yearbook Int Law* 33(38):1–37
- Rauschmayer F, Paavola J, Wittmer H (2009) European governance of natural resources and participation in a multi-level context. An editorial. *Environ Policy Governance* 19(3):141–147
- Ravetz JR (1996 [1971]) Scientific knowledge and its social problems. Transaction Publishers, New Brunswick/London
- Rayner S (ed) (1992) Cultural theory and risk analysis. Praeger, Westport
- Renn O (1992) Concepts of risk: a classification. In: Krinsky S, Golding D (eds) *Social theories of risk*. Praeger, Westport, pp 53–79
- Renn O (1998) Three decades of risk research: accomplishments and new challenges. *J Risk Res* 1(1):49–71
- Renn O (2005) White paper on risk governance. Towards an integrative approach (No. White paper no. 1). International Risk Governance Council, Geneva, Switzerland
- Renn O (2008) Risk governance. Coping with uncertainty in a complex world. Earthscan, London
- Renn O, Keil F (2009) Was ist das Systematische an systematischen Risiken? *GAIA Ecol Perspect Sci Soc* 18(2):97–99
- Renn O, Roco M (2006) White paper on nanotechnology risk governance (No. White paper no. 2). International Risk Governance Council
- Renn O, Walker K (2008) Lessons learned: a reassessment of the IRGC framework on risk governance. In: Renn O, Walker K (eds) *The IRGC risk governance framework: concepts and practice*. Springer, Heidelberg/New York, pp 331–367
- Renn O, Klinke A, van Asselt MBA (2011) Coping with complexity, uncertainty and ambiguity in risk governance: a synthesis. *Ambio* 40(2):231–246
- Roca E, Gamboa G, Tàbara JD (2008) Assessing the multidimensionality of coastal erosion risks: public participation and multicriteria analysis in a Mediterranean coastal system. *Risk Anal* 28(2): 399–412
- Rosenau JN (1995) Governance in the 21st century. *Glob Governance* 1(1):13–43
- Rosenau JN, Czempiel E-O (1992) *Governance without government: order and change in world politics*. Cambridge University Press, Cambridge
- Rothstein H (2009) Talking shops or talking Turkey? *Sci Technol Hum Value* 32(5):582–607
- Rotmans J, de Vries HJM (eds) (1997) *Perspectives on global change: the TARGETS approach*. Cambridge University Press, Cambridge
- Schon DA (1983) *The reflective practitioner: how professionals think in action*. Basic Books, USA
- Schwarz M, Thompson M (1990) *Divided we stand. Redefining politics, technology and social choice*. Wheatsheaf, London
- Shapin S, Schaffer S (1985) *Leviathan and the air-pump: Hobbes, Boyle, and the experimental life*. Princeton University Press, Princeton
- Slovic P (1987) Perception of risk. *Science* 236(4799): 280–285
- Slovic P (2000) The perception of risk. Earthscan, London
- Slovic P, Fischhoff B, Lichtenstein S (1982) Why study risk perception. *Risk Anal* 2(2):83–93
- Soneryd L (2007) Deliberations on the unknown the unsensed and the unsayable? Public protests and the development of third-generation mobile phones in Sweden. *Sci Technol Hum Value* 32(3):287–314
- Sparrow MK (2008) *The character of harms: operational challenges in control*. University Press, Cambridge
- Starr C (1969) Social benefit versus technological risk. *Science* 165(899):1232–1238
- Starr C, Whipple C (1980) Risks of risks decisions. *Science* 208:1114–1119
- Stern RE (2000) New approaches to urban governance in Latin America. Paper presented at the seminar IDRC and management of sustainable urban development in Latin America: Lessons learnt and demands for knowledge, 6–7 April, in Montevideo, Uruguay
- Stilgoe J (2007) The co-production of public uncertainty: UK scientific advice on mobile phone health risks. *Public Underst Sci* 16(1):45–61
- Stirling A (1998) Risk at a turning point? *J Risk Res* 1(2):97–109
- Stirling A (2003) Risk, uncertainty and precaution: some instrumental implications from the social sciences. In: Berkout F, Leach M, Scoones I (eds) *Negotiating change*. Edward Elgar, London, pp 33–76
- Stirling A (2004) Opening up or closing down: analysis, participation and power in the social appraisal of

- technology. In: Leach M, Scoones I, Wynne B (eds) Negotiating change. Edward Elgar, London, pp 33–76
- Surdej A, Zurek K (2009) Food safety in Poland: standards, procedures and institutions. In: Vos E, Everson M (eds) Uncertain risks regulated. Routledge-Cavendish, Oxon
- Tesch NS (2000) Uncertain hazards. Environmental activists and scientific proof. Cornell University Press, Ithaca/London
- Thompson M, Ellis R, Wildavsky A (1990) Cultural theory. Westview Press, Boulder
- US-National Research Council of the National Academies (2008) Public participation in environmental assessment and decision making. The National Academies Press, Washington, DC
- van Asselt MBA (2000) Perspectives on uncertainty and risk. The PRIMA approach to decision support. Thesis, Universiteit Maastricht. Kluwer, Dordrecht
- van Asselt MBA (2007) Risk governance: over omgaan met onzekerheden en mogelijke toekomsten (inaugural lecture, in Dutch). Universiteit Maastricht, Maastricht
- van Asselt M, Petersen A (2003) Niet bang voor onzekerheid (No. Voorstudie nr V.01). Raad voor Ruimtelijk, Milieu- en Natuuronderzoek, Den Haag
- van Asselt MBA, Renn O (2011) Risk governance. *J Risk Res* 14(4):431–449
- van Asselt MBA, Rotmans J (2002) Uncertainty in integrated assessment modelling: from positivism to pluralism. *Climatic Change* 54(1–2):75–105
- van Asselt MBA, Vos E (2006) The precautionary principle and the uncertainty paradox. *J Risk Res* 9(4): 313–336
- van Asselt MBA, Vos E (2008) Wrestling with uncertain risks: EU regulation of GMOs and the uncertainty paradox. *J Risk Res* 11(1):281–300
- van Asselt M, Passchier W, Krayer von Krauss M (2009) Uncertainty assessment. An analysis of regulatory science on wireless communication technology, RF EMF and cancer risks (No. Report for the IMBA project, Work Package 1. Epidemiological research & Animal studies (revised version)). Maastricht University, Maastricht
- Van der Sluijs J (1997) Anchoring amid uncertainty. On the management of uncertainties in risk assessment of anthropogenic climate change. (*Houvast zoeken in onzekerheid. Over het omgaan met onzekerheden in risicoanalyse van klimaatverandering door menselijk handelen*). Unpublished PhD-thesis, Universiteit Utrecht, Utrecht
- van Dijk H, van Rongen E, Eggermont G, Lebret E, Bijkter W, Timmermans D (2011) The role of scientific advisory bodies in precaution-based risk governance illustrated with the issue of uncertain health effects of electromagnetic fields. *J Risk Res* 14(4): 451–456
- van Zwanenberg P, Millstone E (2005) BSE: risk, science and governance. Oxford University Press, Oxford
- Versluis E (2003) Enforcement matters: enforcement and compliance of European directives in four member states. Eburon, Delft
- Voedsel en Waren Autoriteit (VWA) (2010) Voorzorg voor voedsel- en productveiligheid: een kijkje in de toekomst. <http://www.vwa.nl/actueel/risicobeoordelingen/bestand/2000272/voorzorg-voor-voedsel-en-productveiligheid-een-kijkje-in-de-toekomst>. Accessed 14 September 2010
- Vos E (1999) Institutional frameworks of community health and safety regulation, committees. Agencies and private bodies. Hart Publishing, Oxford
- Vos E (2000) EU food safety regulation in the aftermath of the BSE crisis. *J Consum Policy* 23:227–255
- Vos E (ed) (2009) The EU regulatory system on food safety: between trust and safety, vol 1. Cavendish/Routledge, London
- Walker WE, Harremoes P, Rotmans J, van der Sluijs JP, van Asselt MBA, Janssen P et al (2003) Defining uncertainty: a conceptual basis for uncertainty management in model-based decision-support. *Integrated Assess* 4(1):5–17
- Walls J, O'Riordan T, Horlick-Jones T, Niewöhner J (2005) The meta-governance of risk and new technologies: GM crops and mobile telephones. *J Risk Res* 8(7–8):635–661
- Weimer M (2010) The regulatory challenges of animal cloning for food – the risks of risk regulation in the European Union. *Europ J Risk Regul* 1(1):31–40
- Wiedemann PM, Femers S (1993) Public participation in waste management decision-making: analysis and management of conflict. *J Hazard Mater* 33: 355–368
- Wiener JB, Hammitt JK, Rogers MD, Sand PH (eds) (2010) The reality of precaution: comparing risk regulation in the United States and Europe. RFF Press/Earthscan, London
- Wildavsky A, Dake K (1990) Theories of risk perception: who fears what and why? *Daedalus* 4:41–60
- Williams RA, Thompson KM (2004) Integrated analysis: combining risk and economic assessments while preserving the separation of powers. *Risk Anal* 24(6):1613–1623
- WRR (Scientific Council for Government Policy) (2010) Uncertain safety: allocating responsibilities for safety (English version). Amsterdam University Press, Amsterdam
- Wynne B (1982) Rationality and ritual. The windscale inquiry and nuclear decision in Britain. The British Society for the History of Science, Chalfont St Giles

- Wynne B (2002a) Risk and environment as legitimatory discourses of technology: reflexivity inside out? *Curr Sociol* 50(3):459–477
- Wynne B (2002b) Seasick on the third wave? Subverting the hegemony of propositionalism: response to Collins & Evans. *Soc Stud Sci* 33(3):401–417
- Zander J (2010) The application of the precautionary principle in practice: comparative dimensions. Cambridge University Press, Cambridge
- Zinn JO, Taylor-Gooby P (2006) Risk as an interdisciplinary research area. In: Taylor-Gooby P, Zinn J (eds) Risk in social science. Oxford University Press, Oxford



# 45 EU Risk Regulation and the Uncertainty Challenge

Marjolein B. A. van Asselt · Ellen Vos

Faculty of Arts and Social Sciences, Maastricht University, Maastricht,  
The Netherlands

<b>Introduction</b> .....	<b>1120</b>
<b>History: Science-Based Risk Regulation in the EU</b> .....	<b>1121</b>
<b>Current Research</b> .....	<b>1122</b>
The Uncertainty Paradox and the Precautionary Principle .....	1122
Uncertainty Intolerance .....	1124
Boundary Work and Plausibility Proofs .....	1127
Tendency to Consider Uncertainty as Equal to Risk .....	1128
Political Deficit: EFSA as the De Facto Risk Manager .....	1129
<b>Further Research</b> .....	<b>1130</b>

**Abstract:** This chapter aims to understand the role of science and knowledge in the regulation of uncertain risks. In such cases, since scientific knowledge is perceived or portrayed to be limited, experts, stakeholders, or the public have or create doubts about the possibility or severity of hazards. At the same time, regulators habitually turn to science and experts in these cases in order to justify their decisions. This “uncertainty paradox” raises important questions about the role of science, knowledge, and experts in uncertain risk regulation. The analysis reveals that the main challenge for EU risk regulation seems to be as to how to break (out of) the uncertainty paradox. The chapter calls for a systematic comparative research of risk regulation regimes in various domains based on an interdisciplinary approach involving legal and policy sciences as well as STS and risk research. It views that this kind of research has the potential to significantly contribute to risk theory, while at the same time raising new issues and new research questions that would require further interdisciplinary research.

## Introduction

---

The importance of dealing with uncertain risks needs hardly to be emphasized in the post-BSE era. The major institutional shortcomings of national and EU decision-making that were revealed by the BSE (or mad cow disease) crisis triggered a crisis of public confidence in both scientific advice and in the management of risks by EU and Member States authorities. It provided a classic illustration of the complex and vital relationship between science and society and of the difficulties stemming from dealing with uncertainty in science. It led both regulators and the general public to become aware of the risks that are intrinsic to the food industry. Moreover, the continuous stretching of the frontiers of science in areas such as biotechnology has raised public anxiety levels even further. Various such crises combined with public anxiety have, in turn, shaped the way in which risks are perceived and subsequently managed, with strong implications for understandings of political accountability, the role of science, and stakeholder participation.

Post-BSE, therefore, the question of how public authorities should deal with risks and uncertainties has become a predominant concern for the EU in its attempt to achieve one of its main objectives, the free movement of goods and the completion and management of the internal market on which products can freely circulate. How to deal responsibly with situations in which there are suspicions that hazards may exist, although (sufficient) scientific or historical evidence for them is lacking, is in fact a key question that can be derived from Ulrich Beck's (Beck 1986, 1997) agenda-setting critique and provoking notion of “organized irresponsibility.” His sociological conclusions have been mirrored by risk scholars (Wynne 1982; Jasanoff 1993; Hager 1995; Ravetz 1996; Klinke and Renn 2002; Jasanoff 2005; Löfstedt 2005). Not surprisingly therefore, prominent risk scholars have indicated that dealing with “uncertain risks” is a major challenge in current societies (e.g., Beck 1986; Jasanoff 1990; Wynne 1995; Nowotny et al. 2001; Ravetz 2001; Harremoës et al. 2002; Löfstedt 2005; Renn 2006). Such uncertain risks concern situations of suspected, potential hazards, which are usually associated with complex causalities, large-scale, long-term and transboundary processes, and which are generally difficult to control and also transcend human sensory capacities.

This chapter aims to address this key challenge and to understand the role of science and knowledge in the regulation of uncertain risks by examining how various actors deal with

science, knowledge, and uncertainty in the field of EU risk regulation. In such cases, since scientific knowledge is perceived or portrayed to be limited, experts, stakeholders or the public have or create doubts about the possibility or severity of hazards (van Asselt 2005). At the same time, regulators habitually turn to science and experts in these cases in order to justify their decisions (Ravetz 1990; Jasanoff and Wynne 1998; Hilgartner 2000; van Asselt 2005; van Asselt et al. 2009). This situation, which we term the “uncertainty paradox” (van Asselt and Vos 2005, 2006), raises important questions about the role of science, knowledge, and experts in uncertain risk regulation. Our analysis seeks to identify the difficulties of dealing with such risks, to detect ways to address these problems, and to set a long-term research agenda.

## History: Science-Based Risk Regulation in the EU

---

Since the outbreak of the 1996 BSE crisis, there has been much debate on the role of science in European regulatory decision-making and in particular on the need to separate risk assessment from risk management. The Medina Ortega Report of the European Parliament found that in 1990–1994, when the disease had reached crisis levels, the European Commission had suffered from poor internal management; decision-making procedures were not transparent; some national interests held too much weight in the decision-making process; and the resulting legislative controls were not effectively implemented. It moreover concluded, *inter alia*, that the relationship between scientific and political decisions of the EU institutions was blurred. Moreover, it held that the EU institutions, the European Commission in particular, had failed to take public health seriously, and that they had attached more importance to the national interests of agriculture and the interests of industry than the protection of public health. The Report was particularly critical of the committee model (especially the committees composed of national representatives), which it found complex, non-transparent, and undemocratic (European Parliament 1997). The Report thus urged for greater transparency, particularly in relation to the conditions of the functioning and contribution of the scientists on the scientific committees and reform of the rules governing the work of these committees to ensure independence and appropriate funding of the scientists, and the publication of debates within committee and of any dissenting opinions.

Without doubt, the BSE crisis underlined that governance of uncertain risks is difficult, and inherently political. The BSE crisis has been considered an example of “organized irresponsibility” (Hajer and Schwarz 2001). We understand “organized irresponsibility” (Beck 1986) as a situation in which society is unprepared for, and unable to adequately deal with, inevitable surprises, negative consequences, and/or long-term impacts associated with uncertain risks, notwithstanding all institutions and procedures in place and the pretense of certainty and control. The sense of being organized and in control further contributes to society’s unpreparedness and inability, as it appeases people’s awareness and alertness. As a consequence, a situation of organized irresponsibility has negative spiral characteristics. Not surprisingly, in the aftermath of the BSE crisis, fargoing reforms were carried out both at European and national level so as to set and/or restructure the various responsibilities.

**Box 45.1 The EU's Regulatory Structure and Decision-Making Procedures**

The European Union, currently composed of 27 Member States, has a unique institutional structure that is laid down in two basic treaties: the Treaty on European Union and the Treaty of the Functioning of the European Union. It has seven different institutions that are allotted different tasks. For the purpose of this chapter, four institutions are of vital importance: the Council of Ministers, the European Parliament, the European Commission, and the European Court of Justice. Most importantly, in the majority of cases, the Council of Ministers and the European Parliament jointly form the European legislator. The Council of Ministers, or often also referred to only as Council, is composed of the national ministers of each Member State; its composition changes according to the topics dealt with (e.g., when dealing with the environment, the Council is composed of national ministers of the environment). The European Parliament is composed of 737 members who are directly elected. The European Commission is generally charged with the task of promoting and defending the European interests. It is composed of 27 commissioners (one per country) and has the right of initiative in terms of legislation. In addition, it is the EU's executive as it implements at the EU level the legislation set by the European Parliament and the Council. The European Court of Justice consists of two main bodies (General Court and the Court of Justice of the EU), is composed of judges, and is to ensure that the EU law is observed in the interpretation and application of the EU treaties.

Issues about uncertain risks are being mainly dealt with by the European Parliament and Council in their role of European legislator through European legislative acts (regulations and directives) and by the Commission as the risk manager that implements and applies these legislative acts. Of importance are also two other bodies that the EU has created in its institutional structure: agencies and committees (comitology). Today more than 30 regulatory agencies have been created at arm's length of the European Commission. They carry out tasks that vary from adopting decisions as regards the European trade mark to advising the European Commission on the safety of a specific product. The European Food Safety Authority (EFSA) is such an agency and has formally only the power to advise the European Commission on all issues of food safety. EFSA is mandated to act as the risk assessor in the European context. Whenever a company wants to apply for a marketing authorization of a food product as required by specific EU legislation or whenever a Member State wants to deviate from EU law, as in some cases is allowed for, EFSA will give a scientific opinion that is transmitted to the European Commission. Before the European Commission is entitled to adopt a decision, it needs to consult a committee composed of national representatives. This committee-based decision-making procedure is called, in Brussels' jargon, comitology and requires the Commission to submit a draft decision to a committee which will give an opinion on whether it agrees or not with the Commission's draft decision. Depending on the type of procedure, such a committee may ultimately be able to block the Commission and prevent the Commission adopting a decision. Before 2011, this in some cases meant that the Council of Ministers could adopt a decision instead of the Commission.

## Current Research

### The Uncertainty Paradox and the Precautionary Principle

The relationship between uncertainty and knowledge is complex (van Asselt 2005). More knowledge may also imply more uncertainty, as more scientific unknowns are highlighted

(compare van Asselt 2000; Levidow et al. 2005). Surely, uncertainty is in this context not equivalent to no knowledge at all. Scientific experts, whether tacitly or explicitly, do possess what we may call “uncertainty information” (van Asselt and Petersen 2003; van Asselt 2005; van Asselt and Vos 2005, 2006; van Asselt et al. 2009; compare Funtowicz and Ravetz 1990; Morgan 2003; Krayer von Kraus 2005), i.e., insights on which uncertainties are (deemed) important, (attributed) sources of uncertainty, whether and how uncertainty is reducible and which interpretations seem valid in view of the current state of knowledge (i.e., the limits to interpretative flexibility). Acknowledgement of the limits of science in providing conclusive evidence, i.e., the impossibility of full certainty, has led to the development of the precautionary principle. At the same time, all legal formulations (see Fischer 2002; very recently Wiener et al. 2010; Zander 2010) of the precautionary principle include what is called a “knowledge condition” (Manson 2002), i.e., the level of proof needed to trigger application (Petersen and van der Zwaan 2003). Although this knowledge condition is often kept vague or ambiguously formulated, the point is that such a knowledge condition implies that lawyers and policy makers appeal to scientists and experts for some kind of plausibility “proof,” which request has the tendency to run down to demanding conclusive evidence on whether something is a risk. Such requested certainty about uncertain risks seems highly incompatible, if not contradictory, to the notion of uncertainty as the core of the precautionary principle, which implies that neither definite proof nor evidence is available. This leads to a paradoxical situation: on the one hand, it is increasingly recognized that science cannot provide decisive evidence on uncertain risks, while on the other hand policy makers and authorities increasingly resort to science for more certainty and providing conclusive evidence (compare Weingart 1999). This particular pattern in risk regulation we refer to as the “uncertainty paradox”: an umbrella term for situations in which uncertainty is present and acknowledged, but the role of science is framed as one of providing certainty (van Asselt and Vos 2005, 2006). In instances of this uncertainty paradox, policy makers and judicial authorities resort to experts for conclusive evidence and definite answers, despite uncertainty precluding both conclusiveness and definitiveness. In uncertainty paradox situations, a very high level of skepticism as to what science can deliver goes hand in hand with a very optimistic level of confidence regarding what science should be able to deliver (Forrester and Hanekamp 2006).

It seems that this paradox is the consequence of getting stuck in the middle of recognizing the meaning of uncertainty. The inherent limits of scientific observations and studies are recognized, which necessitates to abandon the traditional model of decision-making. The precautionary principle is then seen as “tool to compensate” in situations of unavoidable uncertainty. However, it is often not recognized that uncertainty may also erode the traditional positivistic model of knowledge, in which science speaks truth to power. Although uncertainty is recognized, science is still expected to tell the truth about uncertain risks. Strikingly, advocates of the precautionary principle are willing to rethink regulation, but overlook the need to rethink science and its role in regulation. The uncertainty paradox thus raises important questions about the role of science, knowledge, scientists, and knowledge producers in precaution-based regulation of uncertain risks. It seems that this paradox is the consequence of the difficulty to deal with uncertainty and recognizing its meaning.

Our analysis of Case T-13/99, *Pfizer Animal Health SA v. Council* (hereinafter *Pfizer*) case (European Court 2002a; see also European Court 2002b) illustrates how the uncertainty paradox may become manifest in precaution-based regulatory practice of the EU (van Asselt and Vos 2005). This case concerned a ban of the use of four antibiotics including virginiamycin as

additives in animal feeding stuffs by the EU. This ban was subsequently challenged by the company that produced virginiamycin (Pfizer) before the European Court. The scientific experts on a European scientific committee tried to provide a satisfactory plausibility proof, but uncertainty information crept into their considerations. Instead of following the requested “plausibility proof,” the institutions distracted and reinterpreted pieces of uncertainty information in order to construct uncertainty, which was used as a sufficient condition to apply the precautionary principle. In doing so, the institutions implicitly admitted that reasoning about the “plausibility” of uncertain risk involves normative and subjective judgements, on which basis they implicitly considered it legitimate to “redo” the work originally delegated to the experts.

Although the uncertainty paradox put the scientific committee, the Council of the EU and the European Commission in an uncomfortable straitjacket, it is clear that especially the Court was confronted with a deadlock situation. Such a deadlock, we argued, was the consequence of the way the uncertainty paradox was propagated through the regulatory chain. From the beginning, the role of the experts was framed in terms of providing certainty about uncertain risks. The whole regulatory endeavor further stabilized this illusion, with the consequence that the Court was forced to evaluate the merits and validity of scientific claims. In this case, although the Court managed ultimately to produce a ruling, such jurisprudence can easily be used to abuse the precautionary principle (see, e.g., Marchant and Mossman 2004; Forrester and Hanekamp 2006). A series of such cases, we argue, would be detrimental to the whole precautionary endeavor.

Surely, we recognize that it is a tough challenge to legally accommodate uncertainty. This is even more difficult if we take into account that also EU regulation is at least contextualized, if not dictated, by global regulatory regimes as the World Trade Organization (WTO). Even if we would succeed in developing a coherent precautionary regulatory regime that adequately accommodates uncertainty within the EU context, the WTO’s interpretation of “science-based” may reintroduce vicious paradoxes as the uncertainty paradox. Building upon our reflection and analysis, we concluded that the embedment and integration of uncertainty information in global regulatory processes is an important challenge.

Our analysis of three decision-making processes on uncertain risks, pertaining to the import of genetically modified organisms (GMOs), NK603, GT73, and MON863x MON810, confirmed the uncertainty paradox that was reinforced by the interplay of four mechanisms: (1) uncertainty intolerance on the part of both the risk producer, Monsanto, and the risk assessor, the European Food Safety Authority, EFSA, (2) boundary work which enabled EFSA to claim irrelevance and to construct authority claims which served as building blocks in creating plausibility proofs, (3) the tendency to equate uncertainty with risk, which (further) confined risk producers and risk assessors to the role of uncertainty-intolerant producers of plausibility proofs and technocratic provisions that resulted in (4) an even stronger and de facto political, role for EFSA with the consequence that its uncertainty intolerance became a critical and decisive factor in the interplay (van Asselt and Vos 2008).

## Uncertainty Intolerance

---

First we would like to explain that uncertainty intolerance refers to a situation in which uncertainties are not acknowledged, deemed irrelevant, or are simply evaded, instead of

genuinely and systematically investigated (Wynne 2001; van Asselt et al. 2010). Such intolerance is associated with an unwillingness to demand and produce uncertainty information. In the GMO case studies, we observed first that Monsanto displayed uncertainty intolerance in its framing and assessment behavior. Monsanto's safety assessments seemed deliberate efforts to transform any uncertainty about risk into absolute certainty about safety. To this end, they avoided uncertainty in their communication, they defined it away in their reports, and they seemed to attempt to suppress tests that may question the absolute certainty. Second, we observed that Monsanto violated the obligation to disclose all relevant research as it failed to disclose a rat study. Only upon an order of a German court, Monsanto appeared willing to disclose this study, which indicated that MON 863 caused unexplained kidney damage to rats (Greenpeace 2005).

Apart from the fact that the rat study report was held back by Monsanto, we also noted that this rat study report showed that the tested rats had less mineralized kidney tubules than average. Yet, without giving a proper motivation or justification, Monsanto claimed that this finding was of "no biological significance" (p. 11). Hence, the observed effects are dismissed by Monsanto as irrelevant. Van Asselt (2005) observed that in many practices claims of irrelevance enable experts to suggest certainty. Uncertainty intolerant assessors seem inclined to claim irrelevance, while uncertainty tolerant assessors would have investigated the uncertainty and would have sought to share the uncertainty information. Monsanto furthermore stated that the effect cannot be considered "test related" (p. 11). This we find a quite puzzling statement: There is an adverse effect, yet Monsanto claims that this is not caused by the intake of GM maize. It is not explained how this is possible in a control study setup: The only difference between the two groups of rats was that the rats with altered kidneys were fed GM maize. Here, uncertainty intolerant assessment behavior entailed creating a smoke screen with the aim to keep the uncertainty out of sight. It is intriguing that although Monsanto neutralized the uncertainty in the report itself through claims of irrelevance and a smoke screen, it still decided to conceal the rat study report. The combination of these assessment behaviors – claiming irrelevance, creating a smoke screen and suppression of the report – provide further evidence for our evaluation of Monsanto's stance: uncertainty intolerance is both manifest in the framing and the assessment behavior.

In addition to the uncertainty intolerance of Monsanto, we also observed an uncertainty intolerance on the part of EFSA, which is in line with the earlier observations of Levidow et al. (2005): The opinions of EFSA "generally indicate no uncertainty" that might trigger extra risk management measures (p. 270) and they "have framed scientific uncertainties in such a way that they can be resolved by extra information, or can be readily manageable, or can be deemed irrelevant to any risk" (p. 273). The risk assessor partly inherited Monsanto's uncertainty intolerance, as EFSA's risk assessments were in fact meta-reviews of Monsanto's assessment. This arrangement, which is laid down in the relevant legislation, did not only introduce dependency on the willingness of the applicant to disclose all relevant information, but also in terms of framing and the willingness to disclose uncertainty information. EFSA's room of maneuver is determined by Monsanto's framing and information about uncertainties. The behavior of Monsanto with regard to the tests with adverse effects indicates that this dependency is not a theoretical issue, but a practical difficulty. Another illustration is that during the assessment of MON 863 x MON 810 (EFSA 2004a), the applicant had to repeatedly provide EFSA with additional data for a variety of reasons.

We would however like to stress that the uncertainty intolerance of EFSA cannot only be explained by inheritance of uncertainty tolerance from Monsanto. It is noticeable that EFSA

often sides with Monsanto's evaluations and interpretations of data, and even explicitly "agrees with the applicant," without further explanation or critical discussion of uncertainties that might have been overlooked by Monsanto (as it could/should have done in the rat study affair). EFSA framed its assessment activities as risk assessment, but its actual assessment behavior is very consistent with a safety, i.e., an uncertainty intolerant, orientation. Chalmers (2005) characterizes EFSA's style of reasoning as providing "its own corpus of proof" to find no evidence of risk (p. 658).

It is notable that in all cases, the Commission asked EFSA to "consider whether there is any scientific reason to believe" that the placing on the market of the GMOs "is likely to cause any adverse effects on human health and the environment" (EFSA 2003; EFSA 2004a, b). The particular formulation "whether there is any scientific reason to believe" is relevant if we compare it with the closed question (i.e., whether or not the activity constitutes a risk) the Commission asked in the *Pfizer* case (van Asselt and Vos 2005, 2006). It can be argued that the terms of reference to EFSA were more uncertainty tolerant in the light of the following two elements: (1) instead of asking for a decisive answer and proof on whether the risk is a hazard, the Commission asked for indications that hint at adverse effects, and (2) instead of referring to science as the source of absolute truth and certainty, with the notion "to believe" the Commission seemed to accept that science cannot provide certainty about uncertain risks. As a matter of fact, we observed that also in the regulatory documents related to the directives relevant for the regulation of GMOs and in additional guidelines, the Commission expressed openness to uncertainty or even explicitly called for the description and explicit treatment of uncertainty (Levidow et al. 2005). The terms of reference could therefore have been read as invitation to systematically discuss the uncertainties involved and to provide uncertainty information. Interestingly, however, in all three cases, EFSA provided answers phrased in the following terms: the GMO is "as safe as" the conventional counterpart and it is "unlikely to have an adverse effect on human and animal health and (...) the environment." "Unlikely" can be read as "low risk." Combined with the claim "as safe as," it can be argued that the effective message of EFSA's assessment is a plausibility proof, although such nonrisk contention was arguably not requested by the Commission. This production of plausibility proofs can be read as an uncertainty intolerant interpretation of the goal of the assessment. Observations and analyses of experts involved in risk assessment suggest that EFSA's uncertainty intolerance is indeed not unique, but rather widespread among risk assessors institutionalized in the policy domain (Wynne 2001; Stahl and Cimorelli 2005).

Interestingly, Monsanto's safety assessment and EFSA's risk assessment were similar in that they both were uncertainty intolerant. This resemblance led to accusations of bias and relationships with biotech industry, conspiracy theories, etc. Friends of the Earth (2004), for example, argued that EFSA is used by the Commission to force GMOs onto the market. We do not agree with such views. Building upon our analysis, we argue that EFSA was as uncertainty intolerant and safety-oriented as Monsanto, with the consequence that it is understandable that EFSA felt at ease with Monsanto's arguments and that the assessments were very similar. We want to emphasize that we do neither suggest nor argue that EFSA is deliberately defending or advancing Monsanto's stakes. But, because of the shared uncertainty intolerance, their lines of reasoning were very similar. This similarity is not the point of departure, but in part the *effect* of EFSA's attitude toward uncertainty.

## Boundary Work and Plausibility Proofs

Our analysis of the GMO cases also showed that EFSA managed to evade uncertainty, which enabled it to declare GM products to be safe, through boundary work. Boundary work, a notion coined by Gieryn (1983, 1999), is a strategic and purposeful act in which boundaries are drawn between realms, for example, between science and nonscience and between science and politics. Boundary work involves drawing and maintaining contrasts through selective attributions, which effectively demarcate in order to construct “self-evident justification” and “superiority in designated terrains” (Gieryn 1999). Such boundary work appears to be, as literature shows, not just a matter of formal responsibilities, but an ongoing negotiation process on roles and tasks and how these are portrayed to the others (Bal et al. 2002; Hoppe and Huijs 2002; Halffman 2003, 2005; Jasanoff 1990, 2005). In our case studies we observed how EFSA engaged in boundary work. Dynamic boundary work between risk assessment and risk management as well as between science and nonscience facilitated the construction of claims of irrelevance and authority claims, which were used as building blocks in constructing certainty. EFSA explicitly stated to have been requested to consider scientific objections and not to assess nonscientific ones (EFSA 2003, p. 4). It used the constructed boundary between science and nonscience to argue that possible uncertainties about interference with the European environment and regular maize crops are nonscientific concerns. In this way, the scope of the risk assessment could be minimized. In a similar vein, EFSA constructed a boundary between risk assessment and risk management. Such boundary work enabled EFSA to disqualify and dismiss Member States concerns, which could have been read as uncertainty issues (compare Levidow et al. 2005), as “nonscientific” and “issues of risk management.” EFSA’s boundary work is an example of what Jasanoff (2005) refers to as “less transparent, politically significant boundary work” (compare also Cranor 1990). It can furthermore be argued that earlier boundary work in the processes of setting up the regulatory structures, enabled EU policy to define “agribiotechnology as an expert scientific issues (...) kept separate from socio-ethical issues” (Levidow et al. 2005, p. 266) which facilitated EFSA’s boundary work. It is beyond the scope of our expertise to assess the content of EFSA’s claims: We only describe the assessment behavior. Our analysis demonstrates that in instances that could have been read as uncertainty (Member States’ concerns, adverse effects, open questions), EFSA actively evaded uncertainty through boundary work, instead of discussing these uncertainties and exploring whether and how they may matter. Let alone that the possibility of other and “unknown” uncertainties was systematically considered with an open mind (compare Wynne 2001; Levidow et al. 2005).

EFSA’s responses to concerns regarding the risk of cross-contamination illustrate the dynamic nature of boundary work. With NK 603, EFSA refused to consider the possibility that the GMO could contaminate regular maize crops. It argued that the question of contamination was one of risk management, not of risk assessment and thus felt beyond the limits of EFSA’s remit (EFSA 2003). Landfried (1999) (quoted in Levidow et al. 2005, p. 263) argued: “the difficulty of distinguishing between political and technical questions also provides an opportunity to those who might wish to reduce political questions to technical ones.” In our case, issues of uncertainty were recast as political topics. In this way, boundary work helped EFSA to minimize the scope of the assessment. The cross-contamination issue was also raised with regard to GT 73. This time, however, EFSA did engage in what it had previously

argued to be questions of risk management. Concerns regarding contamination were still quickly dismissed with the reasoning that spillage of the GT 73 plant would likely only occur in ports located in industrial areas. Since these industrial areas offer little room for agricultural cultivation, contact with, and contamination of, other plants was portrayed as highly unlikely. Similar reasoning was employed when contamination concerns were raised with regard to MON 810 x MON 863. Also in that case, EFSA appeared more willing to address issues it dismissed in the NK 603 case. Chalmers (2005) argues that in the case of NK 603, EFSA interpreted its role in the narrowest and most formal sense and thus carried out a very minimal risk assessment. We would argue that minimizing the scope of the risk assessment was an effective way to evade uncertainty.

Boundary work enabled EFSA to construct superiority, which facilitated the production of authority claims. EFSA often mobilized “the scientific literature,” but without specific references or they just agreed with some particular findings or conclusions, without providing further justification. Through this *ex cathedra* style, EFSA presented itself as an authoritative voice. “Believe us, we are scientists,” is the implicit message. The constructed authority is then used as an anchor to evade uncertainty. The boundary work brought about self-evident justification for EFSA’s role in deciding what counts as “new science.” The “new”-requirement (compare Chalmers 2005) in combination with the self-acclaimed authority in this domain enabled EFSA to be satisfied with a reference to earlier assessments. In this manner, boundary work facilitated EFSA to produce certainty about non-allergenicity and hence produce a plausibility proof.

## Tendency to Consider Uncertainty as Equal to Risk

Our case-study research has moreover revealed another mechanism contributing to the uncertainty paradox. We noted that various actors tend to equate uncertainty with risks. Although the notion “uncertain risks” points to situations where risks are highly uncertain, this does not mean that *all* risks are highly uncertain and, importantly, not all uncertainties inhibit dangers. Not all uncertainties are by definition relevant or critical for every aspect in a risk assessment. Equating uncertainty with risks implies that any uncertainty is interpreted as a signal of risk.

In our case studies, both the risk producer (Monsanto) and the risk protestors (Greenpeace and Friends of the Earth) tended to equate uncertainty with risk, notwithstanding the fact that they have opposite positions in the GMO debate. Monsanto’s uncertainty intolerance and the associated desire to prove no effect seemed to be grounded in the belief that zero uncertainty would be equal to zero risk so that it would be considered as 100% safe. As a consequence they seem to fear uncertainty, as it threatens the idea of absolute safety.

The risk protestors also tend to equate uncertainty with risk, but in a different way. Their reasoning seems to be based on the belief that uncertainty is equal to risk which would amount to 100% hazard. For example, both Greenpeace and Friends of the Earth tried to discredit EFSA’s opinions by emphasizing that these opinions had not taken into account irregularities in the molecular characterization of a GMO and that these opinions had not been framed “within the context of continuing scientific debate and uncertainty about fundamental issues relating to its conclusions” (Greenpeace 2004, p. 2; Friends of the Earth 2004, pp. 15–16). They used specific uncertainties to discredit EFSA’s risk assessment.

Such equating of uncertainty with risk may be interpreted as a way to politicize uncertainty. Politicization of uncertainty refers to the role of politicians, stakeholders, special interests

groups and/or the public highlighting, amplifying or attenuating uncertainties in order to serve other interests (e.g., Funtowicz and Ravetz 1990; Smithson 1993; Stocking and Holstein 1993; Pollack 2003). Levidow et al. (2005) suggest that in the debates on GMOs such politicization of uncertainty is taking place: “Policy actors play down or emphasize various uncertainties – to challenge evidence of risk or safety, to justify their stances (...), to pursue greater rigour in demonstrating safety, to mediate among conflicting views and/or to delay (...) decisions” (p. 273). That is not the point we want to raise here. We would like to point to the fact that the risk producer tried to avoid uncertainty in order to demonstrate safety, while the risk protestors highlighted uncertainty to demonstrate risk. This tendency to equate uncertainty with risk sustains the uncertainty paradox as it hampers the production and sharing of uncertainty information. Those who adhere to the absolute safety logic (i.e., zero uncertainty = zero risk) do not welcome uncertainty information. On the other hand, actors who consider uncertainty as a sign of risk tend to politicize uncertainty information, which is, according to Frewer et al. (2003), one of the reasons why experts hesitate to communicate uncertainty.

## Political Deficit: EFSA as the De Facto Risk Manager

---

We furthermore observed in our GMO case studies how regulatory provisions allow EFSA, as the risk assessor, to occupy a central position in the decision-making process, which actually leads EFSA's uncertainty intolerance to increase further. In all three cases, regulatory deadlocks occurred because no agreement for or against import was reached in the relevant comitology committees (created to supervise the European Commission) and/or the Council. To avoid procedural standstill, in such situations the relevant legislative provisions ultimately empower the Commission to adopt a decision. Despite resistance of several Member States, the Commission could issue consents for the placing on the market of NK 603 (European Commission 2004) and GT 73 oilseed (European Commission 2005) under Directive 2001/18, closely following EFSA's opinion. Whereas this procedure is designed for extraordinary circumstances, it became the de facto standard operating procedure in these cases of authorization on GMOs and it gave leeway to adopt authorization in line with EFSA's opinion. The Commission's decisions were a matter of rubber-stamping EFSA's plausibility proofs. As a consequence, it seems safe to conclude that EFSA was the de facto risk manager in the NK 603 and the GT 73 cases.

In the MON 810 x MON 863 case, the decision-making process was more complex, because of disagreement within EFSA. In the framework of comitology, Member States could not reach a majority in favor or against authorization, but this time the Commission could not rely on a plausibility proof. Partly due to pending changes in the regulatory regime, the decision-making process was temporarily blocked. In 2005, EFSA issued a new opinion, wherein it succeeded in producing a plausibility proof (EFSA 2005). This opinion was again the basis for a Commission proposal to be discussed in comitology, but again the Member States could not reach a majority, so that the Commission's proposal to authorize was referred back to the Council to decide. As the Council did not manage to adopt a decision in the prescribed period, the Commission was empowered to authorize MON 863 x MON 810 in line with EFSA's plausibility proof (European Commission 2006).

In view of the societal and political controversies in the EU pertaining to genetic modification (e.g., Durant et al. 1998; Gaskell et al. 2000, 2004; Levidow 2001; Wynne 2001;

Jasanoff 2005; Levidow et al. 2000, 2005, 2007; Horlick-Jones et al. 2007; Lee 2008), it is striking that notwithstanding the political controversy about possible authorization of the three GMOs that we investigated, the decisions about the import were ultimately taken in a technocratic manner: the Commission, following EFSA, simply authorized these products. Our analysis confirms Borrás' conclusion (Borrás 2006) that scientific experts continue to have a central, undisputed position in actual EU regulation of GMOs, notwithstanding rhetoric and institutional changes after the BSE crisis. This would mean that there is a clear political deficit and a trend of scientification of politics (Everson and Vos 2009). It is exactly such technocratic division of responsibility, and associated credibility and legitimacy questions, which were the main political concern after the BSE crisis and the *raison d'être* of EFSA (Vos 1999, 2000; Vos and Wendler 2006). Our analysis adds another dimension to the old concerns: Technocratic provisions sustain or even reinforce the uncertainty paradox in situations of an uncertainty intolerant risk assessor (van Asselt et al. 2010).

## Further Research

---

Over the past years, our interdisciplinary law–social sciences research has sought to deepen and further understanding of the complexities of the governance of uncertain risks and as such to provide input for theory development in the field of risk governance. To this end we have examined various cases to understand how various actors deal with science, knowledge, and uncertainty in the field of EU risk regulation. We aim to provide some insights into the ways in which these actors wrestle with uncertainty. Our analysis suggests that uncertainty intolerance is a core problem. Recognizing and addressing uncertainty can inform knowledge-generation (Levidow et al. 2005). To counteract uncertainty intolerance, uncertainty training is needed. In current education programs, science is presented as a body of certainty (Pollack 2003; see also Collins 1987), which nourishes uncertainty intolerance. In all scientific education, more explicit attention should be paid to uncertainty aspects of science (van Asselt and Petersen 2003). Our own experience with uncertainty training suggests that it is possible to change the attitude toward uncertainty and creates openness for communicating uncertainty, without hampering the willingness to bear responsibility and take decisions. Another aim of uncertainty training is to facilitate awareness of uncertainty paradox mechanisms among people involved in risk regulation. Furthermore, risk assessment bodies, such as EFSA, should not only include experts on issues relevant with regard to specific uncertain risks, but should also welcome uncertainty tolerant experts who are aware of mechanisms associated with the uncertainty paradox in order to organize resistance to the production of plausibility proofs. In addition to this, it might prove fruitful to organize realms where risk producers, risk assessors, risk managers, and risk protestors meet, next to the indirect communication through formal reports. Such two-way exchanges might help to discuss uncertainties in a different way, and might enable risk regulators to gain an understanding of what science can and cannot provide. Notwithstanding the fact that currently we see some ways to counteract uncertainty intolerance, we think that further research is still needed to better understand uncertainty intolerance both on the individual and collective level. Such insights are needed to better target training and to develop strategies to change arrangements and incentives that favor uncertainty intolerance.

On the basis of our research, we feel confident to conclude that the uncertainty paradox is deeply ingrained in current risk regulation arrangements and the broad sociopolitical order.

The main challenge for risk governance seems therefore to be how to break (out of) the uncertainty paradox and to rethink the role of science and expertise in risk regulation (compare Wynne 2001; Levidow et al. 2005; Jasanoff 2005). Insights into, and broader awareness of mechanisms that bring about, and sustain, the uncertainty paradox are only first, but necessary steps. When looking at the current debate on EU GMO regulation, we observe that most critics continue to frame EFSA's role in terms of the uncertainty paradox: EFSA is expected resolve diverging scientific opinions and is still cast as “a body responsible for supplying *unimpeachable* scientific advice and guidance” (Randall 2006, p. 413, emphasis added). It is not recognized that the problems arise from the great burden placed on science as the basis for decisions (Levidow et al. 2005). In this way, an artificial certainty that glosses over any uncertainty regarding risks of GMOs is asked for, which only encourages and reinforces “the self-delusions of institutional science” (Wynne 2001, p. 457) and the “lack of reflexivity about the quality of the knowledge it provides” (Wynne 2001, p. 458). This direction further reifies the uncertainty paradox. Surely, we must stress that in the context of GMOs, the new approach for GMO cultivation that the Commission launched in 2010 addresses the political sensitivities of Member States that do not want to cultivate GMOs on their territory that we also observed in our case studies and recognizes explicitly the need to balance between maintaining the EU system of authorizations based on scientific assessment of health and environmental risks and the freedom of Member States to address specific national, regional or local issues raised by the cultivation of GMOs (European Commission 2010; see Poli 2010). On the other hand, it may be wondered whether this is motivated by the desire to shatter the uncertainty paradox as it does not seem coincidental that such an approach would free the EU institutions from being challenged before the WTO courts and confer unambiguously upon Member States themselves the responsibility to justify possible trade barriers before the WTO. The obligations of the WTO agreements seem indeed to reinforce the need for further study and understanding of EU risk regulation and the role of the various actors and provisions herein, also at the global level.

Recent research on EU regulation on nanotechnology seems to suggest that the uncertainty paradox is present also in this area. Although in this area the regulatory provisions expressly include ethical and other concerns as concerns to be taken into account in the decision-making, Maria Lee concludes that in practice this has

- ▶ not led to the inclusion of ethical concerns in the chemicals or cosmetics regulation, and even in respect of food, decisions are framed largely around risk and safety. These institutional innovations (carving out a space for “risk management”, public participation, expertise in areas other than risk assessment) begin to express the breadth of the decisions to be taken on nanotechnology and other emerging technologies (Lee 2010, p. 820).

She underlines that such innovations “fail to grapple with the basic issue, which is that conventionally we regulate only on the basis of risk. Superficial changes go nowhere when the decision is so heavily constrained by the regulatory context, dominated by a narrow approach to risk assessment. We are faced with a very serious failure of accountability: either decisions are being attributed to science and risk when actually they are political; or the issues that decision-makers say they are taking into account are simply deemed irrelevant” (Lee 2010).

So the next step is to more systematically compare risk regulation regimes in various domains (see Fig. 45.1). Most analyses, including ours, tend to concentrate on a particular technology, while studies systematically comparing risk regulation regimes are lacking. Such

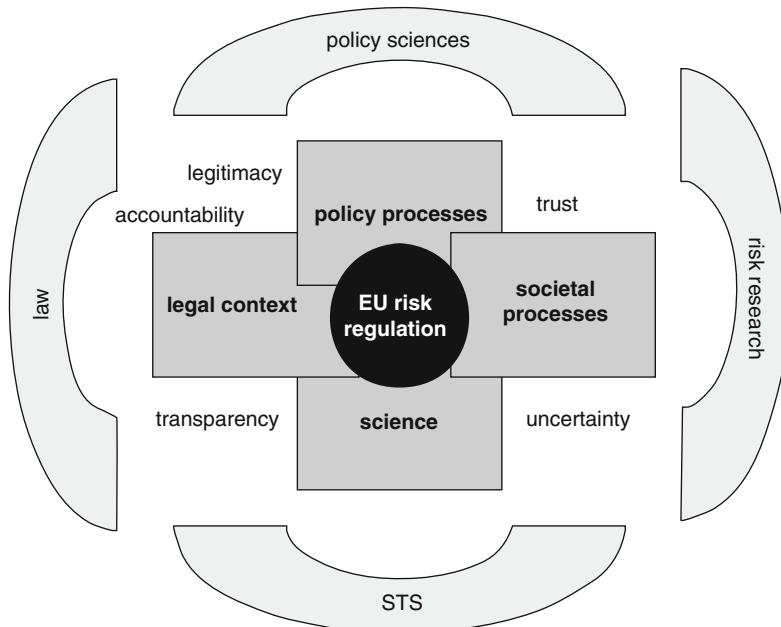


Fig. 45.1

An interdisciplinary research agenda for EU risk regulation

comparative research would enable to examine which problematic patterns seem to be widespread across domains, while differences in dealing with uncertainty might provide a basis for developing strategies to deal with uncertainty in responsible ways. We view four closely intertwined topics of particular importance: (1) the role of science and expertise in policy and decision-making; (2) the question of how to deal with uncertainty and trust; (3) the role of the precautionary principle; and (4) stakeholder participation.

Surely, a central topic will be the examining and rethinking of the role of science and expertise in policy- and decision-making pertaining to innovation, trade, and uncertain risks. In our current research, we have indicated that there is a need to rethink the current arrangements, provisions, expectations, and working styles. In the upcoming years, we hope that further case studies will provide insights into questions such as how and which legal requirements shape the role of science in the decision-making context and what features of science-based decision-making can or may enhance or decrease societal and political support.

As a second theme, we would like to point to the study of the issue of trust. We already highlighted that in situations of uncertainty, trust in the decision-making process is of particular importance. At the same time, due to major regulatory scandals, the current societal climate in which risk governance takes place has been qualified as post-trust (Löfstedt 2005). As far as we know, research examining the relationship between ways to deal with uncertainty and levels of trust is currently lacking. It seems therefore time to thematize the relation between uncertainty and trust.

A third theme on our research agenda is the study of the precautionary principle. Although much research has already been done in this area (see very recent Wiener et al. 2010; Zander 2010), more research seems necessary as regards the precise scope for precautionary action allowed by the trade rules of the WTO. In this context, we view further research and synthesis as necessary as to the role of the precautionary principle in case law and in societal debates. Such research is needed not just to evaluate what is adequate and responsible use of the principle but also to establish what is needed to endorse such best practices.

A fourth theme concerns stakeholder participation. Many writings on risk governance suggest the inclusion of stakeholders as an important step forward (Dreyer and Renn 2009; van Asselt and Renn 2011). However, whether and how stakeholder participation contributes to or rather obstructs decision-making that involves scientific uncertainty has not been investigated yet. It is for example known (van Asselt and Petersen 2003) that stakeholders may also stimulate a quest for certainty. It seems therefore necessary to critically examine how and in which contexts stakeholder participation could contribute to responsible governance of innovation in view of uncertain risks.

This agenda would in our view need to look into specific cases as well as comparative studies based on an interdisciplinary approach. These cases will lead to new empirical insights that will provide a further basis for an empirically informed theory development. On the other hand, also theoretically informed empirical research is needed to test claim pertaining to the uncertainty paradox, uncertainty intolerance, boundary work, and politicization of uncertainty. Although these notions in our view facilitate a more theoretical perspective on risk regulation, as any theoretical proposition they need to be tested. We argue that interdisciplinary research is indispensable to get a better understanding of the problems at stake and possible solutions. Lawyers, policy scientists, and risk scholars tend to put “science” in a black box and take knowledge claims as well as claims about independent expertise for granted, which is problematic in situations characterized by scientific uncertainty. The STS view is needed in order to appreciate what we emphasized above, the “less transparent, politically significant boundary work” (Jasanoff 2005). Social sciences help to question the often “utopic” view on the role of experts, putting them on a pedestal and considering scientific advice as simple facts, as the social science discipline teaches how science is made. Social sciences thus help the legal discipline in the thinking about the legal position of experts. It also helps to understand why certain cases of multiple complexity such as GM authorization may lead to complex regulatory processes but do not resolve the political dispute. The legal view is needed to understand the legal context that frames what actors can (and cannot) and should (and may not) do. The legal science helps for example to get a more realistic view on participation, as the legal discipline requires answers to questions such as who should participate, where and when, and thus plays a part in the thinking about process design and the shaping of participation. Policy sciences are needed in order to understand the policy processes in which regulators operate, which also frames the regulators’ room for maneuver. The risk perspective is needed to understand the kind of policy dossiers regulators deal with. Risk research provides insight into the specific requirements pertaining to risk dossiers, such as trust, participation of stakeholders (from innovators to protestors), and societal processes which may arise around risk (such as social amplification and attenuation of risk). While issues of accountability, legitimacy, democracy, and transparency addressed in legal and policy sciences refer to societal actors in the abstract, both in STS and risk research, the role of societal actors in real situations is explicitly considered. Interdisciplinary law–social sciences research has already yielded new insights

into risk regulation, both on an empirical and theoretical level. This type of research has therefore the potential to significantly contribute to risk theory. At the same time, this type of research also will raise new issues and new research questions that require further interdisciplinary research.

## References

---

- Bal R, Bijker WE, Hendriks R (2002) Paradox of scientific authority: on the societal impact of the health council's advice (in Dutch). *Gezondheidsraad*, Den Haag
- Beck U (1986) *Risk society: towards a new modernity*. Sage, London
- Beck U (1997) The world as risk society. Essay on the ecological crisis and the politics of progress (in Dutch). De Balie, Amsterdam
- Borrás S (2006) Legitimate governance of risk at the EU level? The case of genetically modified organisms. *Technol Forecast Soc* 73:61–75
- Chalmers D (2005) Risk, anxiety and the European mediation of the politics of life. *Eur Law Rev* 30:649–675
- Collins HM (1987) Certainty and the public understanding of science: science on television. *Soc Stud Sci* 17:689–713
- Cranor CF (1990) Some moral issues in risk assessment. *Ethics* 101:123–143
- Dreyer M, Renn O (eds) (2009) *Food safety governance: integrating science, precaution and public involvement*. Springer, Berlin/Heidelberg
- Durant J, Bauer MW, Gaskell G (1998) *Biotechnology in the public sphere: A European source book*. Cromwell Press, London
- EFSA (2003) Opinion of the scientific panel on genetically modified organisms on a request from the commission related to the safety of foods and food ingredients derived from herbicide-tolerant genetically modified maize NK603, for which a request for placing on the market was submitted under Article 4 of the Novel Food Regulation (EC) No 258/97 by Monsanto. *EFSA J* 9:1–14
- EFSA (2004a) Opinion of the scientific panel on genetically modified organisms on a request from the commission related to the notification (reference C/DE/02/9) for the placing on the market of insect protected genetically modified maize MON 863 and MON 863 x MON 810, for import and processing, under Part C of Directive 2001/18/EC from Monsanto. *EFSA J* 49:1–25
- EFSA (2004b) Opinion of the Scientific Panel on Genetically Modified Organisms on a request from the Commission related to the notification (reference C/NL/98/11) for the placing on the market of glyphosatetolerant oilseed rape event GT73, for import and processing, under Part C of Directive 2001/18/EC from Monsanto. *EFSA J* 29:1–19
- EFSA (2005) Opinion of the Scientific Panel on Genetically Modified Organisms on a request from the Commission related to the notification (reference C/DE/02/9) for the placing on the market of insect-protected genetically modified maize MON 863 x MON 810, for import and processing, under Part C of Directive 2001/18/EC from Monsanto. *EFSA J* 251:122
- European Parliament (1997) Final BSE inquiry report. Rapporteur Manuel Medina Ortega. A4-0020/97/A
- European Court (2002a) Case T-13/99, Pfizer Animal Health SA v Council ECR 2002 Page II-03305
- European Court (2002b) Case T-70/99, Alpharma Inc. v Council of the European Union [2002] ECR II-03495
- European Commission (2004) European Commission Decision 2004/643/EC. OJ 2004 L295/13
- European Commission (2005) European Commission Decision 2005/635/EC. OJ 2005 L228/11
- European Commission (2006) European Commission Decision 2006/47/EC. OJ 2006 L26/17
- European Commission (2010) COM(2010) 375 final, proposal for a regulation of the European parliament and of the council amending directive 2001/18/EC as regards the possibility for the Member States to restrict or prohibit the cultivation of GMOs in their territory
- Everson M, Vos EIL (eds) (2009) *Uncertain risks regulated*. Routledge/Cavendish Publishing, London
- Fischer E (2002) Precaution, precaution everywhere: developing a “common understanding” of the precautionary principle in the European Union. *Maastricht J Eur Comp Law* 9:7–28
- Forrester I, Hanekamp JC (2006) Precaution, science and jurisprudence: a test case. *J Risk Res* 9:297–311
- Frewer LJ, Hunt S, Brennan M, Kuznesof S, Ness M, Ritson C (2003) The views of scientific experts on how the public conceptualize uncertainty. *J Risk Res* 2:75–85
- Friends of the Earth (2004) Throwing caution to the wind: a review of the European food safety authority and its work on genetically modified foods and crops. <http://www.foeeurope.org/GMOs/publications/EFSAreport.pdf>

- Funtowicz SO, Ravetz JR (1990) Uncertainty and quality in science for policy. Kluwer, Dordrecht
- Gaskell G, Allum N, Bauer M, Durant J, Allansdottir A, Bonfadelli H, Boy D, de Cheveigné S, Fjaestad B, Gutteling JM, Hampel J, Jelsøe E, Correia Jesuino J, Kohring M, Kronberger N, Midden C, Nielsen TH, Przestalski A, Rusanen T, Sakellaris G, Torgersen H, Twardowski T, Wagner W (2000) Biotechnology and the European public. *Nat Biotechnol* 18:935–938
- Gaskell G, Allum N, Wagner W, Kronberger N, Torgersen H, Hampel J, Bardes J (2004) GM foods and the misperception of risk perception. *Risk Anal* 24:185–194
- Gieryn TF (1983) Boundary-work and the demarcation of science from non-science: strains and interests in professional ideologies of scientists. *Am Sociol Rev* 48:781–795
- Gieryn TF (1999) Cultural boundaries of science: credibility on the line. University of Chicago Press, Chicago
- Greenpeace (2004) Greenpeace critique of Monsanto's roundup ready oilseed rape, GT73. <http://eu.greenpeace.org/downloads/gmo/GPTechCritiqueOfEFSAOpinion.pdf>. Accessed 1 July 2005
- Greenpeace (2005) Monsanto ordered to make secret study public. <http://eu.greenpeace.org/downloads/gmo/MON863GermanCourt050620.pdf>. Accessed 1 July 2005
- Hajer M (1995) The politics of environmental discourse. Ecological modernization and the policy process. Clarendon, Oxford
- Hajer M, Schwarz M (2001) Conturen van de risicomaatschappij. Inleiding. In: Beck U (ed) De wereld als risicomaatschappij. De Balie, Amderstam, pp 7–22
- Halfman W (2003) Boundaries of regulatory science: eco/toxicology and aquatic hazards of chemicals in the US, England, and the Netherlands, 1970–1995. Universiteit van Amsterdam, Amsterdam
- Halfman W (2005) Science-policy boundaries: national styles? *Sci Public Pol* 32:457–467
- Harremoës P, Gee D, MacGarvin M, Stirling A, Keys J, Wynne B, Guedes Vaz S (2002) The precautionary principle in the 20th century: late lessons from early warnings. Earthscan, London
- Hilgartner S (2000) Science on stage: expert advice as public drama. Stanford University Press, Stanford, USA
- Hoppe R, Huijs S (2002) Boundary work between science and policy: paradoxes and dilemma's (in Dutch). RMNO, Den Haag
- Horlick-Jones T, Walls J, Rowe G, Pidgeon N, Poortinga W, Murdock G, O'Riordan T (2007) The GM debate: risk, politics and public engagement. Routledge, London/New York
- Jasanoff S (1990) The fifth branch: science advisers as policy makers. Harvard University Press, Cambridge
- Jasanoff S (1993) Bridging the two cultures of risk analysis. *Risk Anal* 13:123–129
- Jasanoff S, Wynne B (1998) Science and decision-making. In: Rayner S, Malone EL (eds) Human Choice and climate change. Battelle Press, Washington, DC
- Jasanoff S (2005) Designs on nature: science and democracy in Europe and the United States. Princeton University Press, Princeton
- Klinke A, Renn O (2002) A new approach to risk evaluation and management: risk-based, precaution-based, and discourse-based strategies. *Risk Anal* 22:1071–1094
- Landfried C (1999) The European regulation of biotechnology by polycentric governance. In: Joerges C, Vos E (eds) EU committees: social regulation, law and politics. Hart Publishers, Portland, OR, pp 173–194
- Lee M (2008) EU regulation of GMOs: law, decision-making and new technology. Edward Elgar, Cheltenham
- Lee M (2010) Risk and beyond: EU regulation of nanotechnology. *Eur Law Rev* 35:799–821
- Levidow L (2001) Genetically modified crops: what transboundary harmonization in Europe? In: Linerooth-Bayer J, Löfstedt R, Sjöstedt G (eds) Transboundary risk management. Earthscan, London, pp 59–90
- Levidow L, Carr S, Wield D (2000) Genetically modified crops in the European Union: regulatory conflicts as precautionary opportunities. *J Risk Res* 3:189–208
- Levidow L, Carr S, Wield D (2005) European Union regulation of agri-biotechnology: precautionary links between science, expertise and policy. *Sci Public Pol* 32:261–276
- Levidow L, Murphy J, Carr S (2007) Recasting "substantial equivalence": transatlantic governance of GM food. *Sci Technol Hum Val* 32:26–64
- Löfstedt RE (2005) Risk management in post-trust societies. Palgrave Macmillan, Hampshire/New York
- Manson NA (2002) Formulating the precautionary principle. *Environ Ethics* 24:263–274
- Marchant GE, Mossman KL (2004) Arbitrary and capricious: the precautionary principle in the European Union Courts. AEI Press, Washington
- Morgan MG (2003) Characterizing and dealing with uncertainty: insights from the integrated assessment of climate change. *Integr Assess* 4:46–55
- Nowotny H, Scott P, Gibbons M (2001) Re-thinking science: knowledge and the public in an age of uncertainty. Polity Press in association with Blackwell Publishers, Cambridge
- Petersen AC, van der Zwaan BCC (2003) The precautionary principle: (Un)certainties about species loss. In:

- van der Zwaan BCC, Petersen AC (eds) *Sharing the planet: population - consumption- species. Science and ethics for a sustainable and equitable world.* Eburon Academic, Delft
- Poli S (2010) The commission's new approach to the cultivation of genetically modified organisms. *EJRR* 4–2010 Symposium on the EU's GMO Reform, p 339
- Pollack HN (2003) *Uncertain science... Uncertain world.* Cambridge University Press, Cambridge
- Randall E (2006) Not that soft or informal: a response to Eberlein and Grande's account of regulatory governance in the EU with special reference to the European Food Safety Authority (EFSA). *J Eur Public Pol* 13:402–419
- Ravetz JR (1990) *The merger of knowledge with power.* Mansell, London
- Ravetz JR (1996) *Scientific knowledge and its social problems.* Transaction Publishers, Brunswick/London
- Ravetz JR (2001) Models of risk: an exploration. In: Hisschemöller M, Hoppe R, Dunn WN, Ravetz JR (eds) *Knowledge, power and participation in environmental policy analysis.* Transaction Books, New Brunswick/London, pp 471–492
- Renn O (2006) Risk governance: towards an integrative approach. IRGC White Paper No 1, International Risk Governance Council (IRGC), Geneva
- Smithson M (1993) Ignorance and science: dilemmas, perspectives, and prospects. *Knowl Creat Diffus Utilization* 15:133–156
- Stahl CA, Cimorelli AJ (2005) How much uncertainty is too much and how do we know? A case example of the assessment of ozone monitoring network options. *Risk Anal* 25:1109–1120
- Stocking SH, Holstein LW (1993) Constructing and reconstructing scientific ignorance: ignorance claims in science and journalism. *Knowl Creat Diffus Utilization* 15:186–210
- van Asselt MBA (2000) Perspectives on uncertainty and risk: the PRIMA approach to decision support. Kluwer, Dordrecht
- van Asselt MBA (2005) The complex significance of uncertainty in a risk era: logics, manners and strategies in use. *Int J Risk Assess Manag* 5:125–158
- van Asselt MBA, Petersen AC (2003) Not afraid of uncertainty (in Dutch). Lemma/RMNO, Den Haag
- van Asselt MBA, Renn O (2011) Risk governance. *J Risk Res* 14:431–449 (forthcoming)
- van Asselt MBA, Vos EIL (2005) The precautionary principle in times of intermingled uncertainty and risk: some regulatory complexities. *Water Sci Technol* 52:35–41
- van Asselt MBA, Vos EIL (2006) The precautionary principle and the uncertainty paradox. *J Risk Res* 9:313–336
- van Asselt MBA, Vos EIL (2008) Wrestling with uncertain risks: EU regulation of GMOs and the uncertainty paradox. *J Risk Res* 11:281–300
- van Asselt MBA, Vos EIL, Rooijackers B (2009) Science, knowledge and uncertainty. In: Everson M, Vos EIL (eds) *Uncertain risks regulated.* Routledge/Cavendish Publishing, London, pp 359–388
- van Asselt MBA, Vos EIL, Fox T (2010) Regulating technologies and the uncertainty paradox. In: Goodwin M, Koops BJ, Leener R (eds) *Dimensions of technology regulation.* Wolf Legal, Nijmegen, pp 259–284
- Krayer von Kraus MP (2005) Uncertainty in policy relevant sciences. Ph.D. Thesis. Technical University of Denmark, Copenhagen
- Vos E (1999) Institutional frameworks of community health and safety regulation: committees, agencies and private bodies. Hart, Oxford/Portland
- Vos E (2000) EU food safety regulation in the aftermath of the BSE crisis. *J Consum pol* 23:227–255
- Vos E, Wendler F (2006) Food safety regulation at the EU level. In: Vos E, Wendler F (eds) *Food safety regulation in Europe. A comparative institutional analysis.* Intersentia, Antwerp, pp 65–138
- Weingart P (1999) Scientific expertise and political accountability: paradoxes of science in politics. *Sci Public Pol* 26:151–161
- Wiener JB, Rogers MD, Hammitt JK, Sand PH (eds) (2010) *The reality of precaution. Comparing risk regulation in the United States and Europe.* RFF Press, Washington
- Wynne B (1982) *Rationality and ritual: the wind-scale inquiry and nuclear decisions in Britain.* British Society for the History of Science, Chalfont St. Giles
- Wynne B (1995) Technology assessment and reflexive social learning: observations from the risk field. In: Rip A, Misa TJ, Schot J (eds) *Managing technology in society: the approach of constructive technology assessment.* Pinter Publishers, London and New York
- Wynne B (2001) Creating public alienation: expert cultures of risk and ethics. *Sci Cult* 10:445–481
- Zander J (2010) *The application of the precautionary principle in practice: comparative dimensions.* Cambridge University Press, Cambridge

# 46 Risk Management in Technocracy

Val Dusek

University of New Hampshire, Durham, NH, USA

<b>Introduction .....</b>	<b>1138</b>
<b>History .....</b>	<b>1139</b>
Philosophical Concepts and Historical Bases of Technocratic Thought .....	1139
Twentieth Century Western Technocracy .....	1142
Twentieth Century Critics of Technocratic Rationality .....	1144
The German Romantic Idealist Reaction Against Technical Rationality .....	1144
<b>Current Research .....</b>	<b>1148</b>
Two Contemporary Students of Critical Theory: Giving More Credit to Technocratic Reason .....	1148
Technocratic Science Versus “Mythic and Magical Thinking” .....	1150
Technocracy Versus the Magical and Occult Attitude .....	1151
The “Science Wars” and Technocratic Risk Management .....	1152
Psychology of Risk and Technocracy .....	1153
Citizen Investigations of Risk and “Local Knowledge” .....	1155
The Introduction of Technocratic Attitudes and Tendencies in Some Apparently Non-Technocratic Treatments of Risk .....	1155
The Public’s Problem in Identifying Technical Experts .....	1156
<b>Further Research .....</b>	<b>1158</b>

**Abstract:** Technocracy is an idea that has a long history. The idea of rule by experts goes back to ancient Greece, while the idea of rule by scientists or engineers is several centuries old. Formal mathematics, empirical studies, and natural scientific approaches are primary. A recent version of technocracy claims that the technicians do not literally rule but frame the choices of the leaders of industry and government. Technocracy can also be an attitude toward society or a tendency within individual institutions. Technocratic risk management is that version of risk management that emphasizes purely scientifically determined, objective measures of risk and ignores, downplays, or attempts to discredit public fears or estimates of risk that do not correspond to the scientific probabilities of death, disease, and injury. For a technocrat risk communication is either purely descriptive information or an attempt to undermine public fears or concerns with risk that differ from those of the specialists. Technocratic risk management contrasts with democratic or participatory risk evaluation. Despite the fact that no full technocracy exists, technocratic elements may be present in various forms of deliberative democracy and risk management involving public or citizen participation.

## Introduction

---

The term “technocracy” means rule by technocrats or technical experts. Technocracy is similar etymologically to “plutocracy” (rule by the wealthy) and “democracy” (rule by the common people or *demos*). Some people misidentify technocracy with importance of or rule by technical devices. Indeed artificial intelligence, in its more ambitious projected versions, would allow technocracy as rule by technical experts to transform into rule by computers.

The ideal of risk management in technocracy would be for experts in risk management to make objective, scientific risk analyses, and have the results of their analyses acted upon by government and citizens. Citizens would be ordered or persuaded to accept the results of the risk analysis experts and reject their own opinions and intuitions insofar as they disagreed with those of the experts.

Certainly contemporary industrial societies are far from manifesting the ideal form of risk management in a technocracy. The reports of risk management experts are not always unbiased, but sometimes or often influenced by political and economic interests. The opinions of the experts have been challenged by many organizations and publics in recent decades. Widespread distrust of scientific expertise has grown. Conflicting claims of expertise are sometimes made.

Nevertheless, the great number of government committees, commissions, and corporate departments that issue risk assessments and attempt to manage risks, the frequent appeal to risk analyses in public debate, and the number of policies that depend upon risk analyses and risk management show a strong but not all-powerful tendency in contemporary society. Further, even if the society does not always or often function as a full technocracy, the technocratic *attitude* is present in many evaluations and communications of risk. This attitude is one that the only relevant estimates of risk are objective, scientific ones, that the experts know best, that the public ought to trust the statements of experts and follow the directives of the experts.

## History

### Philosophical Concepts and Historical Bases of Technocratic Thought

Two aspects of risk management are relevant to technocracy: (1) The rationality involved in traditional risk analysis is the formal and instrumental rationality that is advocated by and characteristic of technocratic thought. (2) The role mathematical analysis and empirical surveys used by experts, which often is opposed to and contradicted by ordinary popular opinion and intuition concerning risks, where expert scientific opinion trumps popular beliefs is characteristic of technocracy.

Although the term technocracy was occasionally used in the 1890s with various meanings, it was not until the 1920s that the term became widely used and the notion became widely known. An engineer, William Henry Smyth, coined the word in close to its later sense in 1919 (Bell 1973, p. 349 n8). The rise of the Progressive Movement in the USA in the early twentieth century, with its demands for reform and for “social engineering” an engineering-like control of social institutions gave rise to stronger demands for a full technocracy around the time of World War I and the succeeding two decades.

Although significant used of the term technocracy is only a century old, the general notion of rule by experts goes back to ancient Greece and the notion of rule by specifically technological experts goes back four centuries.

Both Socrates and Plato in the fifth and fourth century BCE called for rule by experts. Socrates was highly critical of Athenian democracy, which chose many officers by lot. He argued that just as one would go to a medical expert for cures, to a skilled carpenter to have a house built, or a skilled navigator to pilot a ship, so one should go to an expert in political rule for guidance of the state. Plato in the *Republic* sketched out an ideal state that was ruled by philosopher kings trained in mathematics. Plato was influenced by the Pythagoreans both in his belief that mathematical knowledge was the exemplar of rational knowledge and that rulers should be trained in mathematics. The followers of Pythagoras claimed that the world is made of numbers and that numbers are the key to knowledge. The Pythagorean movement combined this philosophical claim with a religious and political movement whose adepts learned secrets of number theory (Von Fritz 1977). In the *Republic* (Plato [c 380 BCE] 1992) the rulers studied higher mathematics (including musical theory and astronomy) for 10 years before they became apprentice administrators. However, for the earlier Plato, mathematics only provided training in mathematics as abstract, logical reasoning as a preliminary for training in philosophy. In the *Republic* dialectic, or philosophical reasoning, is higher than mathematical knowledge. However, toward the end of his life, Plato in the *Philebus* (Plato [c 355 BCE] 1975) wrote of a “science of normative measure,” a kind of mathematical ethics, perhaps developed as an answer to the growing influence of the leading mathematician Eudoxus, who taught a kind of utilitarian ethics which was gaining supporters in Plato’s academy. Plato also began to discuss “ideal numbers,” which were higher and more abstract than ordinary numbers. According to a report of the music theorist Aristoxenus, Plato, toward the end of his life, announced a public “lecture on the good.” The audience who expected uplifting moral edification instead found the lecture focusing on higher mathematics and becoming more and more difficult. The majority of the audience left, in boredom and dismay (Aristoxenus and Pearson 1990; Findlay 1974; Gaiser 1980; Kramer 1990). Speucippus and Xenocrates, Plato’s immediate

successors as second and third heads of the Academy, moved toward identifying all abstract forms with numbers, whether ordinary numbers for Speucippus or ideal numbers for Xenocrates, for whom even the soul was a number. It was these views that Plato's greatest student, Aristotle identified with the Plato's Academy and rejected.

The two sides of Platonic doctrine of concern here, the identification of mathematics with a higher form of knowledge and the notion of rule by experts, were revived but were for the most part separately during the early modern period of the seventeenth century. On the one hand, the so-called rationalists, René Descartes, Gottfried Leibniz, and Baruch Spinoza developed the notion of mathematical reasoning as the highest form of reasoning and a model for philosophy. Descartes used his constructive analytic geometry, which he had discovered, as a key to and model for knowledge. Leibniz, who invented an early mathematical logic, identified God's knowledge and creation of the universe with a kind of logical reasoning, surveying all possible worlds, and based on the principles of noncontradiction and identity. Leibniz thought that he could develop a "universal characteristic," a mathematical logical language representing the essential concepts in all thought, and that could settle all disputes by simple logical deductions (Leibniz [1680] 1966). Leibniz claimed that in the future, if there was a dispute in any area, the opponents would say simply, "Come. Let us calculate." Leibniz also used the principle of continuity, based on the differential calculus that he also invented, as a model for the range of objects in the universe (Leibniz [1680] 2001). Spinoza, though not a supreme mathematical inventor as were Descartes and Leibniz, used the logic structure of Euclid's geometry, with theorem and proof, as his model for exposition in the *Ethics* (Spinoza [1677] 2000).

The other strand of early modern thought contributing to later technocracy is Francis Bacon.

Bacon in England in the earlier part of the seventeenth century contributed to the notion of technological and scientific rationality as the source of later British empiricism. He was the main proponent and propagator of the inductive method in science. Bacon was also a very early proponent of the role of scientific advisors in increasing the wealth and welfare of the nation. Bacon presented an ideal society, "The New Atlantis" in which inventors, engineers, and what would later be called scientists, advised the kingdom and increased the wealth, health, and welfare of the state through their studies of nature, chemical experiments, and development of machinery. Bacon's advisors in "Salomon's House" did not literally rule, but functioned as high-level advisors. In some cases they kept secret from the ruler some devices that they considered too dangerous (Bacon [1624] 1989).

Rather surprisingly, Bacon, in strongest contrast to the European rationalists, rejected the importance of mathematics (Quinton 1980; Jardine 1973; Dijksterhuis 1961). While both Descartes and Bacon ridiculed the traditional logic of Aristotle and the medieval schoolmen, Descartes replaced it with geometrical construction, and Bacon replaced it with empirical induction. Bacon's contempt for mathematics as useless formalism like the quibbles of the medieval professors, and the hypothetical nature of the Copernican system led him to reject the Copernican theory of the sun as center of the solar system as not empirical (Jardine 1973).

Later followers, such as Robert Boyle and Isaac Newton, combined the inductive, experimental method with mathematics.

Only in the nineteenth century did the two components of Plato's original notion of rule by experts become reunited. The French social thinkers and philosophers, Henri de Saint Simon and August Comte, hangers-on at the engineering school, the *École Polytechnique*, envisioned

a society ruled by experts in physics, engineering, and the social sciences in the modern sense. St. Simon, who coined the terms “physicist” and “industrialism” among many others, envisioned a “Council of Newton” composed of scientists, engineers, and other experts who would rule and plan society. After the reaction to Napoleon and the monarchist restoration, St. Simon included bankers in a place of honor among his experts (Saint-Simon 1952; Manuel 1962). Bankers, like physicists, dealt with numbers. St. Simon’s shifts concerning what professions would be represented in the ruling council foreshadow twentieth century disagreements within technocratic thought of the place of natural scientists versus social scientists as the ruling technocrats.

August Comte, unlike St. Simon, was actually enrolled at the *École* for several years, studying the leading mathematical physicists of the day, such as Carnot, Laplace, and Lagrange, but was expelled for his antimonarchist political views. Comte literally wished to replace religion with science and envisaged a hierarchical society similar to the Catholic Church, but with scientists as high priests, and perhaps himself as sort of positivist Pope. Comte coined the name sociology for what he considered the master science.

Comte was not only a major figure in the prehistory of technocracy, but the major figure in the origin of the doctrine of positivism. Comte entitled his major works with the name “positive,” positive philosophy and positive polity. Positive for Comte was natural scientific knowledge. Comte interpreted the natural sciences as purely predictive, not explanatory. He claimed that knowledge (and with it society) evolved through “three stages,” the theological, the metaphysical, and the positive. In the first stage, religious explanations of phenomena are given, gods, spirits, and demons, gradually evolving unity of God from fetishism (with spirits in everything) to polytheism, to monotheism. In the second stage, the metaphysical, explanations are given in terms of powers and essences. Finally, in the positive stage, laws of sequence are used to predict. A science that claims to explain the essence of things is still partially mired in the metaphysical stage. Comte’s view, less elaborately presented and without its historical sequence account is presented in much contemporary scientism. Science is seen as the highest form of knowledge.

For the early twentieth century logical positivists, empirical science is really the only form of knowledge. Ethics, religion, and metaphysics are all cognitively meaningless. The attitudes of Comtean positivism and logical positivism live on among those today that see science as the highest or the only form of genuine knowledge. Likewise those who disparage the claims of the humanities to offer some sort of valuable understanding of the world are heirs of the positivist tradition (Comte [1830] 1988). The logical positivists also emphasized the sharp separation between facts and values, taken over by technocratic theories, with facts objective and values subjective, noncognitive emotional expressions.

The logical positivists or Vienna Circle of the 1930s combined the positivistic thesis that science was the highest, indeed the only, form of genuine knowledge, with the mathematical logic that had been developed by Bertrand Russell a few decades earlier. They made the linguistic and logical foundations of the empiricist approach to science far more rigorous than had the earlier positivists, claiming that statements for which empirical verification was impossible were meaningless. The logical positivists dropped the Comte’s historical thesis concerning the evolution of knowledge and Comte’s detailed planning of a social utopia. Many of the logical positivists originally combined their scientific worldview with support of social democratic politics. However, upon fleeing Nazi Germany and Austria for the USA as immigrants, and residing in America as somewhat insecure foreigners during the McCarthy

anticommunist campaign, dropped any explicit social doctrines and retreated to “the icy slopes of logic” (Reisch 2005).

Technocratic tendencies in St. Simon were absorbed into Marxism and Soviet Communism by the fact that Karl Marx’s collaborator Friedrich Engels incorporated many phrases and much terminology of St. Simon in his expositions of Marxism. Engels’ presentations were much easier to read than those of Marx and had a great influence on the socialist and communist movements. Thus Lenin and Stalin focused on phrases in the writings of Engels which actually originated with St. Simon, for instance “society as one great factory,” “administration of things and not of men,” “artists as engineers of the soul” (Hayek 1955). Soviet Marxism had a strong technocratic trend, partly from earlier non-Marxist Russian utopians, but also from the St. Simonian strain in Marxism (Bailes 1978).

## Twentieth Century Western Technocracy

---

During the first three decades of the twentieth century engineers professionalized. Particularly in USA the engineering professions issued official statements concerning the importance of engineers for the progress and prosperity of society. By the 1920s the claims of these professional manifestos reached their height. Engineers were portrayed as the paradigm of objective and rational thinkers (Layton 1971, p. 63). Most working engineers shied away from the actual technocracy organizations and parties, but their vision of their role was quite similar to that of the militant technocratic campaigners.

The major figure in American technocracy was the economist Thorstein Veblen (brother of leading geometer, Oswald Veblen). Veblen contrasted the “predatory spirit” of businessmen with the “instinct for workmanship” of the engineers. He saw business practices as mainly wasteful and engineering as tending to quality products and efficient use of resources (Veblen [1921] 1983). Veblen at the height of the Russian Revolution even once proposed “a soviet of engineers” to rule society (Dowd 1964). Veblen’s ideas fit with the Progressive Movement’s emphasis on “social engineering” and efficiency. Veblen’s “institutional economics” influenced many in President Franklin D. Roosevelt’s “brains trust” was much as did the ideas of John Maynard Keynes. Stuart Chase, a follower of institutionalism and efficiency theory apparently coined the term “New Deal,” used to designate Roosevelt’s program.

In contrast with this indirect influence on the New Deal, the political organization that called itself Technocracy, Inc., after a brief spurt of popularity in the early Great Depression, collapsed in ridicule into a small fringe organization after the very poor public performance of its leader, a self-styled engineer who turned out to lack a degree (Elsner 1967). He turned out to be a charismatic “coffee house engineer” like the many self-proclaimed but untalented coffee-house poets. Technocracy as an organization survived into the new millennium, but only as a tiny, virtually unknown organization.

Despite the virtual disappearance of the openly political Technocracy organizations, technocracy as an idea or trend continued to influence thought during the later twentieth century in Europe and the USA. Gunnar Myrdal, the Swedish economist was a major proponent of technocratic planning (Myrdal 1960). Karl Mannheim, once he moved to Britain from Hungary, defended an important role for intellectuals in “democratic planning,” although Mannheim’s idea of the intellectual or social scientist was not that of the pure technician or pure empiricist.

The Manhattan Project during World War II in the USA, led by the physicist J. Robert Oppenheimer, led to a leading role for nuclear physicists as policy advisors due to the highly technical and secret nature of the relevant physics and engineering of nuclear weapons and missiles. After Oppenheimer's fall from political grace during the Cold War nuclear arms race, other leading physicists with more strongly anticommunist views such as Edward Teller, father of the H-bomb, and John von Neumann became major government advisors (Bird and Sherwin 2005; Goodchild 2004; Heims 1980; Herken 2002). The rise of game theory (also stemming from von Neumann) led to the rise of nuclear strategy experts, including figures such as Bernard Brody, Hermann Kahn, and Daniel Ellsberg (Kaplan 1983; Poundstone 1992). It was not really until the 1960s that the defense intellectuals and researchers at the RAND corporation were fully invited into the Defense Department by Robert S. McNamara of the "Whiz Kids," a name for his analysts that had previously been applied to the post-World War II Ford Motor Company executives, including McNamara.

During the 1960s and 1970s a number of writers in sociology such as Daniel Bell in the USA and Ralph Dahrendorf (1967) in Germany portrayed modern politics as a matter of what Karl Popper called "piecemeal social engineering." Bell proclaimed *The End of Ideology* (1960) in the title of one book, claiming that ideological fanaticism had been replaced by social scientific tinkering and pragmatism. Economists, apparently successful during the boom of the 1950s and 1960s with Keynesian methods to fine-tune the economy likewise proclaimed the replacement of traditional political ideologies with planning by experts. A group of social scientists, including Daniel Bell (1973) and Zbigniew Brzezinski (1970) propounded the importance of scientific and technical experts (including social scientists like themselves) in managing society.

The theory of technocracy was tied to the theory of postindustrial society. It was claimed that industrial society centered on manufacturing was steered by capitalists, but that postindustrial society with its emphasis on service industries, knowledge and information would be ruled by technocrats. Not all advocates of the postindustrial society thesis were technocrats, but most of the late twentieth-century social science technocrats were postindustrial theorists.

A more subtle form of technocracy was propounded by the institutional economist John Kenneth Galbraith (strongly influenced by the writings of Veblen). According to Galbraith, the stratum of technical experts, such as economists, accountants, engineers, and managers, which he called the "technostructure," do not literally rule society, as do the scientist-priests or engineer-planners in the utopias of St. Simon, Comte, and Veblen, but rather frame the choices and supply the information on the basis of which corporate CEOs, military generals, or politicians make their decisions. Thus, according to Galbraith's account, corporations and governments of industrial societies are guided by the technostructure although the decisions are officially made by the leaders, who in some cases are little more than figureheads or the public face of the organization. This sense of technocracy may be relevant to government risk analysts. In the financial industry, some hedge fund leaders are themselves financial risk analysts by training, while in other firms the risk analysts are advisors in the technostructure. Galbraith's claims about how the technocratic managerial firm has replaced traditional capitalism (Galbraith 1967), downgrading the importance of profits for long-run stability, have been criticized effectively both by economists and by the developments of subsequent decades. Insofar as Galbraith's noncapitalist managers, more concerned with technological rationality than with profits were replaced by initiators of capitalist buyouts and managers whose remuneration consists in stock options, the notion that the top managers are technocrats

and not concerned primarily with profits has been largely discredited. Nevertheless, at levels below the top managers, risk analysts and risk managers important are part of the techno-structure in Galbraith's sense and have gained increased importance in government agencies and private corporations since Galbraith presented his thesis.

## Twentieth Century Critics of Technocratic Rationality

---

The main forms of technocratic rationality stem from the legacy of on the one hand Plato and the seventeenth century rationalists, and on the other hand Bacon and the British empiricists. Both of these forms of rationality are used and valued in technocratic risk analysis. Highly formalistic mathematical models in risk analysis of technological projects and even more so in financial risk analysis are valued and pursued. Empirical evidence from surveys and physico-chemical and biomedical experimental analyses are also central to modern risk management.

A partial list of procedural goals of technocratic rationality includes analysis, planning and precise calculation, as well as impersonal, standardized, and precise criteria (Gendron 1977, p. 58).

The formal mathematics and logic valued by the rationalists, Descartes, Leibniz, and Spinoza discussed above is combined with emphasis on observation and empirical evidence emphasized by the British empiricist followers of Bacon such as John Locke and John Stuart Mill. Risk benefit analysis itself mirrors the structure of the utilitarian ethics of Jeremy Bentham. Utilitarianism is a form of consequentialism. Only consequences, not motivations, intentions, or principles, count for evaluating actions and policies, in contrast to Christian ethics of intentions or Kant's ethics of principles. Risk calculations are made of the positive and negative consequences of project for all parties affected. While Bentham calculated results in terms of pleasure and pain, risk benefit analysis generally calculates results in terms of money. The early mathematical micro-economics of Francis Y. Edgeworth and Stanley Jevons is based on utilitarian ethics of satisfaction. Even today a few economics texts start with measure in terms of "utils" (units of satisfaction or utility in Bentham's sense) before transferring to monetary measures. Thus there is a natural, historical link between microeconomics and utilitarian ethics and of both with risk benefit analysis. With few exceptions (and these very tentative ones such as that of Cranor), consequentialist ethics, or ethics that evaluates acts or policies in terms of causal consequences, is implicitly embraced by risk analysts. Indeed risk analysts explicitly speak of analysis of consequences (Rechard 1999). A number of the traditional problems for risk benefit analysis, such as those of justice, and human dignity or worth mirror or parallel problems for Bentham's utilitarianism. Competing theories of ethics, such as Christian ethics, Kantian ethics, or various ethics that emphasize intentions, as well as intuitionist ethics, that rejects the possibility of formal calculations in ethics, or even of formal rules for ethics, are generally ignored by risk analysts while working within their specialty. The few recent attempts to develop non-consequentialist risk analysis are sketchy and preliminary (Cranor 2007).

## The German Romantic Idealist Reaction Against Technical Rationality

---

Within German philosophy from the late eighteenth century Romantic Movement to the present there have been several tendencies of thought that reject the technocratic model of reason such as formalistic rationalism and British empiricism. This model would include

the methods of quantitative risk analysis. The appeal purely to empirical data concerning risk and purely mathematical analyses and calculations concerning risk are seen as involving an inadequate model of reason. The alternative, more informal and philosophical, notion of reason is an outgrowth of the reflections in the late eighteenth and early nineteenth century on dialectics as a form of reason higher than calculation or technical reasoning. In the twentieth century, particularly through so-called Critical Theory, these conceptions have influenced philosophy in the rest of the world. Immanuel Kant ([1781] 1996), despite his valuing the certainty and universality of mathematics and the fundamental principles of theoretical physics, also dealt with those areas wherein human reason runs, or attempts to run, beyond the bounds of reasoning dealing with common sense objects or with the objects of science. Kant called this sort of reason about everyday and scientific objects “understanding” (a term that acquired a very different meaning in the social science of the late nineteenth century). Kant described understanding as dealing with objects as delimited, bounded entities in space and time. Kant’s “Aesthetic” (dealing with sensory experience) dealt with the most fundamental forms of sensory perception, space and time, while his “Analytic” dealt with the fundamental categories that govern thought both commonsensical and science about objects. However, Kant, in his “dialectic” dealt with the tendency of reason to attempt to go beyond the bounds of sense in metaphysics and to deal with purported entities which were not objects in the being delimited, finite, and bounded or in space and time. The prime examples of these non-objects or pseudo-substances which reason wished to treat as objects are the universe as a whole, the soul, and God. In treatment of the universe as a whole, understanding attempts to extrapolate from its reasoning about finite spatiotemporal objects to the infinite. This leads to antinomies in which it possible to disprove claims both that the universe is finite and that it infinite in space and similarly to disprove both that the universe is eternal and that it has an origin in time. (These antinomies resemble and may have influenced the early understanding of the antinomies or paradoxes of mathematical naïve set theory.) In the case of the soul, understanding runs beyond its bounds in attempting to treat an unobservable, non-spatiotemporal object as a sort of physical object and runs into logical fallacies or paralogisms. In the case of God, understanding attempts to deal with something both infinite in numerous attributes (omniscient and omnipotent) and non-spatiotemporal, leading to fallacious proofs of the existence of God. If one takes the “low” view of Kant’s reason, it simply understands run amuck, spinning its wheels, venturing beyond its allotted capabilities. For Kant mere “thinking” is the use of the forms of understanding without any empirical or sensory input.

Kant himself had attempted to limit the claims to unlimited metaphysical knowledge by the rationalists; however, his successors, claiming to follow Kant, nevertheless reinstated the claim that total knowledge is possible. Philosophers after Kant, such as Friedrich Schelling and Georg W. Hegel, took a “high” view of Kant’s Reason. Kant had also considered reason the faculty of the critique of knowledge itself as well as the setter of ultimate goals for the unification of knowledge. Schelling turned Kant’s reason into a sort of aesthetic faculty of intuition, possessed in the highest form by geniuses, reversing Kant’s claims as to the centrality of discursive, sequential thought and his claims about the modest powers of the understanding. Hegel took Kant’s dialectic, which had been a kind of logic of illusion, and made it into a positive means of gaining knowledge of the infinite. Hegel claimed that for reason to know its supposed limits it had in some sense to pass beyond them. For Hegel, like the earlier rationalists, the infinite is in a sense prior to the finite. Hegel then claimed that dialectical reason could grasp the universe and its development in totality. Furthermore, since for Hegel reality is “spirit” the dialectic is

not just a means of reasoning but the dynamic and structure of reality itself. Marx took over this notion of the dialectic but denied that reality is spirit, claiming it is concrete natural reality. These notions of dialectical reason were used by the Frankfurt School to contrast with technocratic reason and to offer an alternative to purely calculative or purely means-end reasoning.

One of the major schools criticizing technocracy and the empirical/formal rational notion of reason in the twentieth century was the Frankfurt School of Critical Theory. Max Horkheimer (1972), Theodor Adorno (Horkheimer et al. [1948] 2002), and Herbert Marcuse (1964) took over Hegel's notion of Reason and Hegelian version of the Marxist dialectic to criticize technocratic reason. Horkheimer and Marcuse reviewed numerous works of logical positivism in an almost wholly negative manner, seeing positivism as reflecting the rise of technocratic reason and the decline of critical reason. They contrasted Hegelian, critical reason, with "instrumental reason," the means/end reason of technology and social tactics. The sociologist Max Weber et al. ([1914] 1968) had claimed that instrumental action involving means and ends cannot determine ultimate ends. For Weber the ultimate ends cannot be rationally justified. They are the outcome of an irrational, existential decision. Weber claimed that the main trend of modern, western societies was rationalization of all fields of activity (Weber [1904] 2001) (Weber et al. [1914] 1968). Bureaucracy, with its strict roles and chains of command, was a crucial form of this rationalization of society. Nevertheless the ultimate goals were based on irrational choices, and social solidarity was maintained by irrational charisma. There has been only limited application of Weberian rationalization theory to environmental issues (Murphy 1994).

Horkheimer and Marcuse claim, in contrast, that there can be reasoning about ultimate values. The claim of sociologist Max Weber and the existentialist philosophers that values cannot be rationally justified, but must be the result of a nonrational decision or leap, as well as the claim of the logical positivists that value statements are "cognitively meaningless" are both contributors to the "eclipse of reason." Critical theorists see these trends as contributing to what Adorno calls "the administered world," a world totally controlled by technocratic planners, in which popular opposition or critical objections by intellectuals are suppressed or co-opted. Although the first generation critical theorists did not discuss and apparently were not aware of risk analysis, they would have seen technocratic risk analysis as a prime example of this technocratic administration, in which the decisions by technical experts overrule popular opinion and humanistic claims. The elimination of questions of ends and values from rational discussion allows the implicit values of those in control to go unchallenged.

Marcuse goes so far as to claim that formal and instrumental rationality should be replaced with dialectical rationality. He hints at the replacement of present science and technology by a new "liberated" version, perhaps influenced by Marx's phrase "the liberation of nature" and Ernst Bloch's utopian and occultist development of the conception. These hints are not followed up, and most likely could not be by Marcuse. Indeed, in his later writings, Marcuse sees natural science used in a free and cooperative society as liberating, and not in need of such a total transformation.

Jürgen Habermas of the "second generation" of critical theory in a lengthy article on Marcuse entitled "Technology and Science as Ideology" (Habermas 1970) qualifies Marcuse's earlier identification of science and technology as such as oppressive and as means of social control. Habermas claims that the problem is not, as it was for Horkheimer and the Marcuse of *One Dimensional Man* (1964) instrumental reason, or physical science and technology as such

but rather the application of such means-end reasoning and orientation toward control to the social and human sciences. For Habermas, scientism is the misapplication of natural science and technological methods to the human sciences (Habermas 1973).

Habermas, who is highly eclectic, incorporates the distinction between the natural and human sciences that was developed in late nineteenth century German historical and social thought. This tradition was in strong opposition to positivism, both in its earlier, Comtean, version, and in its later logical positivist version. Habermas's lines of criticism of positivism and the misuse of or overreaching by science often called scientism and his linking these tendencies to technocratic rule via the authority of mathematics and physical science clearly apply to quantitative risk analysis as a means to technocratic authority.

Late nineteenth century neo-Kantians of the Southwest German School distinguished between the natural science and human sciences. Wilhelm Windelband distinguished nomothetic (lawful) sciences from idiographic (individual) sciences. According to Windelband, history and cultural sciences focus on unique individuals (are idiographic), while the natural sciences focus on laws (are nomothetic). Heinrich Rickert added to this the claim that natural sciences are value-free while the human sciences incorporate values and are value-oriented (Rickert [1896–1902] 1986, [1899] 1962). Much of the late twentieth century psychological studies of risk tend to treat humans as rational decision makers whose choices follow a formal calculus, and treat human behavior along the lines of the physical sciences. Also many technocratic risk analysts, such as the early Starr and Whipple (Starr and Whipple 1980), profess their own total value-neutrality.

Another distinction made by Wilhelm Dilthey is between natural sciences that deal with causes and the human sciences that deal with meanings (Dilthey et al. [1883] 1989). According to Dilthey the human sciences use “understanding” as opposed to causal explanation. “Understanding” (*Verstehen*, in a very different sense from Kant’s use) involves rethinking or reliving the experience of others, a kind of empathetic participation (although the roles of intellectual rethinking and empathetic refeeling were debated in later discussions of *verstehen*). The sociologist Max Weber attempted to combine *verstehen* in understanding the values and goals of actors and cultures with a causal analysis of social and historical events, nevertheless keeping the two methods separate.

Habermas, in his contrasting the (appropriate) instrumental reason of causal analysis in the natural science with the, for him, inappropriate uses of it in political understanding, contrasts Wilhelm Dilthey’s hermeneutics or interpretation of meaning with the instrumental, natural science approaches of positivism and pragmatism. Hermeneutics was originally the interpretation of biblical texts. Friedrich Schleiermacher in the early nineteenth century generalized it to the interpretation of literary texts in general, and Wilhelm Dilthey at the end of the nineteenth century generalized it to interpretation of human history and culture in general. Hans-Georg Gadamer developed Dilthey’s hermeneutics using concepts from the phenomenologist Martin Heidegger, purging the psychological orientation of Dilthey’s earlier attempts. In contrast to the logical positivists, who advocated the development of a unified science structured deductively and sharing the methods of the physical sciences, the hermeneutic approach claims not only the different hermeneutic method for the human sciences but the philosophical priority of the hermeneutic approach to that of causal or natural scientific analyses. Clearly the contrast between instrumental reason and hermeneutics parallels the contrast between strictly quantitative and utilitarian risk benefit analysis and qualitative and cultural accounts of risks.

Habermas claims that technocracy and scientism in their illicit application of engineering control and instrumental reason to society are ideologies of bureaucratic and administrative society. In his subsequent writings, Habermas contrasts the social system (which can be analyzed by systems theory) with the *lifeworld* which is understood through personal experience. For Habermas creeping technocracy is manifested in the invasion or colonization of the lifeworld by the system, that is, personal experience in face-to-face contacts of humans is superseded by scientific or pseudoscientific accounts based on the approach of the physical sciences. This process includes political and administrative as well as epistemological aspects. Bureaucratic and government regulation of many aspects of personal life through social welfare, educational, and other bureaucracies manage aspects of personal, subjective life. Cass Sunstein's *Nudge* (Thaler and Sunstein 2008) advocating subtle governmental framing of decisions and directing of lifestyle habits and beliefs is a more recent and subtler program for this. Within Sunstein's "libertarian paternalism" agents make choices, but the choices have been framed in such a way that the framers (nudgers) get the choices they want made. The critical theorists would have had a field day with this form of apparent freedom that is guided or manipulated.

Certainly the replacement of the lay public's estimates of risk with the results of quantitative, scientific, empirical estimates of risk by risk managers, and the use of risk communication to shape public attitudes and behavior would be a more recent aspect of this "invasion of the lifeworld by the system" that Habermas decries, whether rightly or wrongly. Habermas discusses the systems theory of the 1960s but does not discuss risk analysis and management as such. The conflict between public perceptions of risk and the scientific estimations of risk fits nicely into the critical theorists' contrast of popular democracy with technological and bureaucratic administration and with Habermas' contrast of system of lifeworld.

Habermas's critique of technocracy and technocratic reason thus grows out of the conceptions of a dialectical reason that supposedly reaches beyond the range and competency of purely technocratic reasoning, with its emphasis on formalism and detached and neutral science.

## Current Research

---

### Two Contemporary Students of Critical Theory: Giving More Credit to Technocratic Reason

---

Two more recent writers on risk trained in critical theory and making intellectual appeals to its texts diverged in various ways from the founders. William Leiss, who began studying under Marcuse and whose first book, *The Domination of Nature* (Leiss 1972) was a survey of theories of technological domination of nature based on critical theory, went on to be heavily involved in the study of risk management in Canada. Leiss even in his early work was critical of Marcuse's utopian hopes for an "emancipatory science," supporting Marcuse's later views that ordinary science deployed by a democratic and cooperative society would help human emancipation. Leiss supports the objectivity of science (Leiss 2001). His studies of the BSE (Bovine Spongiform Encephalopathy) or "mad cow" problem in Britain and Canada with its devastating effects on the cattle industry make him a severe critic of the failed risk management and risk communication, involving initial silence or denial by the British and Canadian governments and the failure to effectively communicate risks until too late (Leiss 2004). Leiss is far more sympathetic to scientific risk analysis than one presumes the early critical

theorists would have been, though he puts far more emphasis on the democratic aspect of risk communication than the technocratic risk managers would. In a number of respects Leiss is more accepting and favorable toward scientific risk analysis than the first-generation critical theorists would have been. Indeed he mentions that Habermas seemed early to suggest that there was more to the critique of scientific rationality than that of scientism and technocracy, but gave up on that theme (Leiss and Chociolko 1994). However, Leiss focuses on the issues of communication and controversy, not on a technocratic pronouncement on scientifically measured risks.

Another figure influenced by critical theory is the sociologist Ulrich Beck. His *Risk Society* (Beck 1992) is an academic best seller in Germany. His own sociological theory of “reflexive modernization” has been contrasted with Habermas’ theory of early modernization and the rise of the public sphere. Beck refers to and used the work of Adorno and Horkheimer. His works are written in an engaging and fluent style, made apt citations of literary works, and claim to present an alternative to other postmodern theories. Striking aphorisms and *obiter dicta* abound. The problem is with the logical coherence of the striking and often insightful theses. Beck’s claim that we are living in a “risk society” early captured the trend toward greater and greater importance for risk analysis in contemporary society and politics. However, he generalizes this notion to the extent of claiming that the distribution of bads (risks) has replaced the distribution of goods of classical capitalism, and that class distinctions are no longer relevant to risk. In this last respect his theory resembles other, competing theories of postmodernism. Classical, competitive capitalism has been replaced by something else. For other postmodernists it is postindustrial society, in the earlier postindustrial theorists of the 1970s with a technocracy, while for the postmodern social theorists of the 1980s and 1990s, especially after the collapse of the USSR with universal neoliberal capitalism. Beck offers the alternative of the risk society. One of his claims is that in past societies risks were in the form of natural disasters, plagues, and wars, but that in modern society the risks are primarily produced by technology itself. This may have appeared true in the period of Chernobyl, Bhopal, and the Challenger disaster, but the first decade of the new millennium, with Indonesian Tsunami, Chinese earthquakes, Icelandic volcano soot, and New Orleans US Katrina hurricane show that natural disasters still have an important place, though often, as in the case of Katrina, or of poorly constructed Chinese school buildings, social and political factors exacerbated the natural disaster.

Beck certainly is among the critics of technocracy and partisans of more democratic treatment of risks. However, his account of science is ambiguous. He does point out correctly several features of the changing role of science. One is that countermovements, particularly in ecology, have their own scientists, such as Rachel Carson and Barry Commoner (Carson 1962). Another is that the hegemony and authority of science is being contested, in part by critiques using science itself. Beck’s description of the society of conflicting experts sponsored by opposing interest groups and institutions as a kind of pluralistic technocracy. Just as classical theories of representative democracy and authoritarian rule were replaced by the theory of pluralism of competing groups, so centralized technocracy in Beck is replaced by a competition and circulation of technocrats, similar to the circulation of elites in Wilfredo Pareto and Gaetano Mosca’s theories of rule by elites or by a “political class” in which social transformations are merely replacements of one elite by another.

Beck states that science has lost its role as a servant of truth. However, he does not spell this out. Some of his theses seem similar to those of the social epistemologist Steve Fuller, in his

book *Science*, that science has become so tied to government and corporate institutions that it resembles more the science of medieval China than it does the independent and largely amateur science of the Europe of centuries past (Fuller 1997). However, Beck does not explicate what the status of science is. It could be purely instrumental, predictive, and manipulative, as the critical theorists (as well as Moritz Schlick the leader of the logical positivist Vienna Circle) claimed it to be. Some of his examples point to the ability of ordinary citizens with their “local knowledge” to outdo technical experts in the discovery of risks. However, he often points out how the countermovements have their own experts and need to put their claims in scientific risk management form in order to gain a hearing from the government agencies. However, his emphasis on the central role of science in society seems not to call for an elimination of technical risk analysis. Beck perceptively captures and incorporates many trends of contemporary society but does not seem to have well-worked proposals for how to handle them (though, neither do most other social theorists).

### Technocratic Science Versus “Mythic and Magical Thinking”

---

Another contrast of scientific and technological rationality is that with so-called “primitive” or “magical” thinking (Malinowski 1954, Wilson 1970). This contrast is sometimes used by defenders of quantitative, scientific risk analysis against popular attitudes toward and evaluations of risk. Popular thought is sometimes castigated as primitive and magical in contrast with the advanced, scientific results of risk analysis. A sophisticated development of this claim and contrast by is the work of Mary Douglas, someone who has achieved great eminence in the study of taboos in indigenous societies, in collaboration of political theorist Aaron Wildavsky, *Risk and Culture: An Essay on the Selection of Technical and Environmental Dangers*. In their cultural theory of risk, Douglas and Wildavsky make the analogy between popular fears of the risks of environmental dangers such as nuclear power, chemical additives, air pollution, electromagnetic radiation, and so forth with noncivilized people’s notion of taboo, uncleanness, and danger. According to Douglas and Wildavsky there is an objective issue of the reality of risks. They claim that environmentalists are not responding to real risks in the environment, but are resisting and resenting authority. They compare environmentalists’ concerns for purity of water and air to indigenous societies’ beliefs about purity and dirt (Douglas and Wildavsky 1982). Douglas had earlier written a highly influential anthropological work, *Purity and Danger: An Analysis of Concepts of Pollution and Taboo*, discussing these doctrines in indigenous African and other cultures that she had studied. While use of the terms “primitive,” “tribal,” and “magical thinking” are often bandied about by opponents of popular environmental fears or critics of drugs or procedures in mainstream medicine, Douglas, as an anthropologist, wishes to claim that magical and mythical thought are part of so-called modern or advanced cultures as well as indigenous ones. However, Douglas in her joint work with Wildavsky disdains and castigates environmentalists clearly as marginal, sectarian, and “primitive,” while treats the judgments of technocrats of the core institutions of government and corporations as rational, not prone to mythical thought. This totally contradicts the perspective of Douglas’ other works, but supports the technocratic understanding of risk management (Douglas 1985).

One problem with contrasting the (correct) scientific-bureaucratic estimates of risk with the (magical or irrational) popular conceptions of risk accounted for by cultural theory is the

issue of the entry of magical thinking into scientific and technological thought as it is actually pursued by real people in institutions as opposed to the ideal forms of scientific and technological thought. Douglas herself admits that residents of modern societies engage in mythic or taboo thought. Can this apply to scientists and risk analysts themselves? This problem is similar to that of the sharp contrast between value-free risk analysis and the value-laden popular conceptions of risk. Just as scientific investigations can incorporate values in several senses, values within scientific method, values involved in worldviews that support scientific paradigms, and values involved in the interests of social institutions or professions, (Mayo and Hollander 1991) so apparently rational and objective scientific thought, and even more so the reception of its authority by the public, including intellectuals not specialists in the relevant science, can manifest aspects of magical thinking. We have seen in the prehistory of technocracy section how the father of positivism, Comte, simply transposed the hierarchy of the Catholic Church with a hierarchy in which the priestly role was taken by scientists.

The authority of scientists, although in recent decades under attack, can have aspects of the reverence for magicians. This was true of nuclear scientists in the early days of the Cold War (Kowarski 1971). Similarly it was true of medical spokespeople. However, this reverence for nuclear physicists and medical doctors in the 1950s has been undermined by the end of the Cold War and various patients' rights movements.

## Technocracy Versus the Magical and Occult Attitude

---

The technocratic trend in risk management contrasts the objectivity, universality, and publicity of science with the secrecy and particularity of prescientific, magical, and mythic thought, often attributing the latter to opponents of technocratic approaches.

The ideal of science is one of universalism, openness to criticism, sharing of data, organized skepticism, and disinterestedness. Karl Popper (Popper 1945, 1962) in philosophy and Robert King Merton (Merton 1947, 1973) in sociology presented these as, respectively, ideal and acting norms of science.

Science with its openness and publicity was traditionally contrasted with the secrecy of magic and alchemy. The very name "occultism" embodies this contrast. However, with the growth of secrets of military research and secrets of corporate research and development this stark contrast of previous centuries between scientific openness and occultist secrecy has been undermined (Kaplan 1983). Writers such as Lewis Mumford (Mumford 1967) have compared the military scientists with the priests of the ancient empires. Alvin Weinberg, head of Oak Ridge Laboratories, is credited with the term "nuclear priesthood" for the engineer custodians of long-lived nuclear waste (Weinberg 1972, Lapp 1965). Similarly a work on the US Federal Reserve and its monetary fine-tuning was entitled "Secrets of the Temple" (Greider 1987).

Risk benefit analyses of technological projects usually make their results, if not their evidence and methods, fully public. This hardly fits with the image of scientist as occult figure. However, quantitative financial risk analysts, so-called quants, did at the opening of the twenty-first century have a kind of prestige partially based on mystery.

The mathematics involved in hedge fund risk analysis (for instance, Wiener processes from Brownian motion, Feynman path integrals from quantum mechanics, and the Ito calculus and its even more complex progeny) is of a sort accessible only to those with graduate educations in math and physics, which most of the quants possess (Lindsay and Shachter 2007; S Patterson

2010). There was a mystique and prestige of advanced mathematics in convincing ordinary investors as well as less mathematically adept bank officials that this sophisticated mathematical apparatus gave assurance of safety. Further, the computer programs used by financial risk analysis “quants,” including high speed trading programs, impressed investors; some were very sophisticated ones with technological complexity and power (Leinweber 2009). In addition, the actual details of the risk formulae and computer programs of most hedge funds are carefully guarded trade secrets, protected by all sorts of computer safeguards from hackers and industrial spies in less competing hedge funds. Of course, events such as the credit crunch in the mortgage securities market of 2007–2008 and the surprising and unexplained “flash crash” of computerized trading in the spring of 2010 demonstrated to many that the claims of mathematical and technological trustworthiness of the formulae and programs were mostly overrated. The complex formulae had assumptions, and a number of the assumptions used were false. The historical memory of the young, ex-physicist quants did not go back to the crash of 1987 let alone to that of 1929. Most of the quants, trained in mathematics, computer programming, the physical sciences, and many of those trained in highly formal economic theory lacked any sense of history, society, or institutions (Skidelsky 2009). Measurements such as value-at-risk (VAR or VaR) turned out to be inaccurate, based on insufficient data from the recent past, and assuming uncertainty can always be probabilistically estimated using mathematical models (Taleb 2010, p. 225n, 2004, p. 289n). Assumptions that different kinds of assets have independent and uncorrelated probabilities turned out to be wrong. In crises many assets’ values showed themselves to be highly interdependent and correlated (Cassidy 2010, pp. 274–278; Triana 2009). The faith of not only naïve individual retirement fund investors but apparently sophisticated European bank officials was based on the mathematical and technological incomprehensibility of the hedge funds’ risk analysis and management apparatus, a kind of occultist faith in complex, arcane, mathematics. Perhaps the greatest harm that the defunding of the Texas Superconducting Supercollider did was not the loss of technological spin-offs, but the driving of unemployed particle physicists into the financial markets, where their impressive but misapplied mathematics mislead even sophisticated institutional investors.

### The “Science Wars” and Technocratic Risk Management

---

The so-called Science Wars in the 1990s (Lingua Franca 2001) were triggered by natural scientists feeling that science no longer had the prestige and respect that it had during the previous few decades. The cancelation of funding for the superconducting supercollider in Texas was symbolic for physicists. Paul Gross and Norman Levitt both defend a somewhat unarticulated positivism, clear mainly in its rejection of social constructivism and theories of the objects of science as subjective. In some respects, the science warriors’ defense of science veers into a defense of scientism and in its denunciation of any who question the pronouncements of technical experts, a kind of technocracy. Gross and Levitt blame the antinuclear movement, the ecology movement, the women’s movement, and Afrocentrism were blamed in part for undermining of faith in science. Paul Gross, author of the chapter on ecology movements in *Higher Superstition*, emphasizes the debunking of claimed environmental risks by citing the example of Alar (Gross and Levitt 1994, p. 163) and being rather skeptical of global warming (p. 158). Sociological studies of the politics of science and literary studies of the rhetoric of scientific papers were seen as undermining the majesty and mystique of science.

The mathematician Norman Levitt, one of the leading and most aggressive polemicists in the science wars castigated Sheila Jasanoff (Jasanoff 1986, 2005) for criticizing establishment risk benefit analyses and supporting popular input into risk management. He notes she supported what he claims is a quack medical movement, clinical ecology, and, according to his own account of unrecorded remarks, creation science in the classroom (Levitt 1999, pp. 222–229). Despite these feelings by certain natural scientists that they were getting no respect, aspects of the scientific mystique remained strong in areas such as genetic explanations of human cultural behavior in sociobiology and evolutionary psychology, widely propagated in the popular media, interest in personal genetic screening, and popular fascination with theoretical physicists' and cosmologists' explanations of the origin of the universe.

## Psychology of Risk and Technocracy

---

The early or classic works on risk analysis were developed by people with an engineering or scientific training. They were mainly developed in reaction to what were seen as excessive, irrational fears of nuclear reactors and nuclear power on the part of the public (Shrader-Frechette 1985a, b). The Electric Power Research Institute sponsored research and education in this area in the USA. Another major motivation for early developments in scientific risk management was what was perceived as irrational fear of pesticides and lab-made chemicals in food. Bruce Ames famously compared the risk of a number of foods and environmental influences (Ames et al. 1987). A famous case not involving nuclear power often appealed to was that of alar-tainted apples in contrast with numerous other foods. This became (though some still dispute the account) a code word for excessive caution concerning tiny risks and costly but unnecessary regulation (Glickman and Gough 1990).

The work of Daniel Kahneman and Amos Tversky (Kahneman et al. 1982; Kahneman and Tversky 1973) concerning the erroneous probability judgments humans make has been used by those with a technocratic bent to disparage popular concerns about risks. Kahneman and Tversky's striking experiments show that humans estimate strongly different probabilities of events, depending how the problem is phrased, in terms of gain versus loss. Human intuitive judgments ignore the conjunction rule in probability and are biased by the base point from which they make estimates. An extreme conclusion from Kahneman and Tversky's work is that of the linguist Noam Chomsky, that although humans have innate and high-functioning linguistic and arithmetic modules, humans lack a probability module (and, according to Chomsky, in some of his speculations, also lack an ethical module).

Even mild or partial technocrats such as Cass Sunstein and Mary Douglas appeal to these results to claim that popular movements concerning nuclear power, food additives and pesticides, and so forth are mistaken and irrational.

It was thought by the early, engineering trained risk management experts that simply informing the public of the correct, scientifically ascertained probabilities would dispel widespread apprehension concerning nuclear plants and pesticides. This proved, of course, to be false. The next tack taken by more technocratically inclined risk managers was to attempt to reassure the public by discussion and to understand the factors other than actual probability of death, illness, or injury that determined the public's fears and estimates of risk. Paul Slovic (Slovic et al. 1981; Slovic 2000) has been a leading figure in estimating these other factors in the public perception of risk.

Involuntary, catastrophic, unknown, novel, long-term, and low-benefit risks were estimated as less risky than voluntarily undertaken, noncatastrophic, well-known, familiar, and high-benefit risks. The more technocratic approach to the psychology of risk then took these other factors into account in terms of dealing with the public, but saw them as mistaken and distorting of the real risks in terms of mortality, morbidity, or injury. The least technocratic risk management policies took these factors into account in evaluating actual designated risks. However, incorporating the psychology of risk (as well as social theories of popular risk estimation) can be used simply as more scientific data concerning objective human behavior to be combined with the natural scientific data concerning physical and biological sources of risk. This social scientific data can then be used to steer the public in the desired direction, either by government directives, commands, and punishments at the authoritarian extreme, or by economic incentives and other “libertarian paternalist” “nudges” lacking explicit coercion, framing of the decision situation by “choice architecture” at the other, less explicitly coercive, extreme.

Gerd Gigerenzer (Gigerenzer 2000) has written extensive criticisms of Kahneman and Tversky’s approach, claiming that humans have “quick and dirty” or rather “fast and frugal” “simple heuristics that make us smart” for rapid decisions that function quite well in real life. He has criticized the omniscient model of an individual with unlimited search time and memory. He claims that his account of human judgment is “rationality for mortals” (Gigerenzer 2008). Although Gigerenzer has written extensively about medical diagnosis and counseling, he has not directly tackled technological or economic risk analyses. Since Gigerenzer’s and colleagues’ work cannot any longer be simply ignored, technocratic risk managers can reply by granting Gigerenzer that our simple decision procedures may have evolved for survival, and that they often function effectively where everyday snap decisions are demanded, but that humans are prone to major errors and fallacies when using these approaches. Kahneman and Tversky among others make a distinction between System I reasoning involving these intuitive “gut” decisions and System II reasoning involving logical deduction and explicit discursive reasoning. Thus some credit is given to our evolved intuitive procedures, but they are discounted as useful in scientific risk estimation. It is claimed that we ought to switch to System II reasoning in dealing with societal risks.

Gigerenzer points out that Kahneman’s own studies show that even professional statisticians who are not prone to the fallacies in textbook, formal probability problems, commit some of the same fallacies as statically untutored when dealing with concrete, real-life problems (Kahneman et al. 1982, pp. 23–31, 46–47). In the same vein, Gigerenzer found that physicians are often almost as bad as patients at interpreting statistical data concerning risks of side effects of medicines or medicines or medical procedures (Gigerenzer 2007). This fact can be used to support anti-technocratic doubts about the infallibility of experts. However, no one, to this author’s knowledge, has advocated using Gigerenzer’s “gut feelings” to deal with technological risk management.

On the other hand, it should be noted that physicians, working engineers on dangerous projects, and financial analysts estimating risks of investments often do appeal to “gut feelings” in making their final decisions, even when scientific and mathematically structured inputs are available. Richard Feynman once claimed that one does not really have full grasp of an equation in physics until one can make intuitive estimates of solutions without doing calculations. There is, of course, a century-old body of writings on the role of intuition and “gut feelings” in judgment in physics and chemistry (Duhem [1914] 1954; Polanyi 1958). This would suggest that even in mathematical, technical-scientific risk analysis System I reasoning has a role.

The intuitive reasoning of mathematicians does not show that the intuitive judgments of risk by the public are trustworthy or accurate. However, it does show that professional risk analysts may use intuitive, unconscious processes in finding solutions to equations and in making judgments concerning what assumptions to accept or what alternatives to include in their analyses.

### Citizen Investigations of Risk and “Local Knowledge”

---

In science and technology studies of traditional non-Western or indigenous science the “local knowledge” of geography, botany, weather, navigation of hunter-gatherers or non-literature peoples in general is contrasted with the “universal” knowledge of natural science that originated in the West. Western science or mainstream modern science is claimed to be universal in several senses, it uses universal laws, unrestricted in space and time for explanation. It is universal in applying to and being pursuable by peoples of all cultures. Science studies people and anthropologists sympathetic to ethno-science or indigenous knowledge have turned this contrast on its head by claiming that modern, mainstream science is itself “local knowledge, not universal knowledge, and its locality is the laboratory” (Harding 1998; Latour 1987). The laboratory can be reconstructed in distant locations, can travel, but it is claimed that laboratory knowledge with its purified substances, frictionless motion in a vacuum, purebred strains of model organisms, itself really local and special, not universal. The local knowledge idea can be applied to conflicts between citizens and scientific risk managers (Jasanoff 1986, 2005). Local citizens may have detailed anecdotal and practical knowledge of terrain, ocean currents, flora, or weather that risk management experts called in from a distant location may lack. Local citizens have identified toxic waste sites denied to exist by corporate experts (Lash et al. 1996). Louisiana fishermen showed justified skepticism concerning claims by representatives and risk communicators of both the BP Corporation and the US government that oil from the Deepwater Horizon Gulf of Mexico oil spill of April 2010 would not reach shore. (The Minerals Management Service’s risk management plan at the time referred to protection of walruses in the subtropical Louisiana waters that even the most ignorant citizen there would know did not exist.)

### The Introduction of Technocratic Attitudes and Tendencies in Some Apparently Non-Technocratic Treatments of Risk

---

The anthropological treatment of risk is at face value non- or anti-technocratic. The role of community tradition, community attitudes and worldviews, belief systems, and myths would seem to go against the Comtean or positivist priority of scientific rationality. It does so in many inquiries such as those of Brian Wynne (Wynne 1982). However, Dame Mary Douglas, despite holding to the universality and centrality of cultural taboos and conceptions of purity and impurity, in her collaboration with Adam Wildavsky, at least, applies the “primitive” or mythic account to the so-called sectarians of the ecology and health movements but not to the “center” of corporate and governmental opinion (Douglas and Wildavsky 1982). Thus this anthropological treatment of risk comes close to the technocrats vision of rational leaders and irrational, pre-civilized public. A consistent application of the anthropological approach can, and often

does, treat the rulers and technocrats themselves as possessing cultural worldviews, traditions, and myths that belie the rational and objective self-image of the technocrats.

Ulrich Beck is supportive of the ecological, health, and other countermovements in contemporary society. He claims that in reflexive modernity the countermovements have their own scientific experts. Although there is a pluralism of expert centers and expert claims, it is in a sense a pluralism of mini-technocratic centers. Although Beck claims that science has given up its claim to be serving truth, it is science to which Beck believes all factions must appeal. The notion of nonscientific, psychological, or anthropological constructions of risk is given little credence.

Third, Cass Sunstein in particular, but some of the other proponents of deliberative democracy as well, give a prominent and even dominant role to the technocrats. Sunstein himself claims that deliberative democracy must have a preponderant role for technocratic opinion. The deliberative side of deliberative democracy allows for time to ponder and evaluate scientific surveys and claims. Sunstein goes further, claiming that the government can act to “distract” the public from fears and opinions that conflict with objective, scientific risk-benefit analysis. He also advocates that the government surreptitiously support “independent” experts to give their opinions and evaluations supporting programs supported by the government. He even advocates the sending of undercover government agents into groups that support irrational conspiracy theories (apparently judged as such by him) to question and undermine these theories. Sometimes the government must undertake wasteful expenditures to eliminate minor risks if the public has great fear of them. This is done on a purely utilitarian consideration that the fears themselves can create great harms in public’s behavior and must be dealt with.

Thus even in a number of contemporary apparently non-technocratic approaches and views, such as the anthropological approach to risk, the self-reflexive critical society of reflexive modernization, and deliberative democracy, technocratic directions and tendencies are present.

## The Public's Problem in Identifying Technical Experts

---

Given that in technocracy the opinions of experts ought to be believed and followed, the problem arises – “Who are the experts?” (Crease and Selinger 2006).

This problem is apparently easy to solve in a totalitarian society, where the experts are simply those designed as such by the rulers or the government. For instance, in Marxist–Leninist societies, dialectical and historical materialism was considered not just a political theory or a philosophy, but a *science* of society and of nature. The ruling party itself acted on this “science” of Marxism–Leninism, and it had the authority to designate the experts in particular fields (Bailes 1978).

However, in a more open society, the problem of identifying the experts is not quite that simple. Certainly there is the whole process of professionalization and credentialing, which sociologists have studied in great detail. An expert is a member of a research institution, university, or government committee which has been accredited by the relevant accrediting bodies. There is a tacit social contract between the professional societies and the society at large. The implicit contract gives the professional body a monopoly on certain services in return for society’s gaining the services of the members of the professional body. The members of the profession are granted a certain self-determination and autonomy (Lowrance 1976). Academic freedom is an extreme example of this professional autonomy, although recently subject to controversy and constraint, and not always followed in the past.

The public often listens to “experts” who are indeed experts in one field of science (and have gained media attention) but who are not particularly expert on the field on which they pronounce in the mass media. An example of this is Steven Hawking’s pronouncements on climate change.

Hawking is a renowned cosmologist and astrophysicist, but his statements on the dangers of climate change are regarded by many in the media and the public as expert on climate because of his skills in relativity theory, black holes, and cosmology, which are unrelated to the issue in question (Connor 2007). Perhaps an even more extreme example of this was the public talks and pronouncements of William Shockley on racial differences in IQ. Shockley was an inventor of the transistor and a brilliant engineer, but his pronouncements on race and IQ were based on no research of his own. Unlike Arthur Jensen or Philip Rushton, who use statistics and surveys to defend their controversial so-called race realism, Shockley defended his position purely by obiter dicta (Shurkin 2008).

Clearly science in a normative sense can differ from science in an institutional or sociological sense at a given time, although retrospectively the poor science may be normatively rejected.

Since the seventh decade of the twentieth century widespread public criticism and opposition to experts has arisen. This was stimulated by the role of scientists, both natural and social, in the Vietnam and other recent wars, the public fears of nuclear power despite expert risk analysts assuring the public of its relative safety, and various health controversies, such as concerning AIDS, and food controversies such as BSE. The genuine expertise of experts has been called into question. Counter-experts, supported by ecology groups or opponents of genetic engineering have appeared. In court the situation is more complex. A variety of experts and alleged experts have been called upon as expert witnesses. In the controversy over recovered memory in child molestation, the expertise of some court-designated “experts” has been strongly denied by other social scientists and writers.

Some, like Ulrich Beck, in his *Risk Society* and Anthony Giddens, claim that with ecology and health movements and controversies scientific expertise has been severely fragmented, and the citizen must decide which of the numerous competing expert views on a topic to follow.

However, being credentialed as an expert and actually following the norms of science can be quite different things. There is a difference between the descriptive, sociological question of whether someone has status in scientific or technological professions as social institutions, and the normative issue of whether some member group within those institutions are following rationally legitimate or valid procedures. Sir Cyril Burt is an extreme example of the possibility of dichotomy between public reputation or authority and genuinely scientific quality of research. Burt was highly respected in his profession of psychometrics. He edited the most prestigious journal in the field, *The British Journal of Statistical Psychology*, advised the London Council on the tracking of students, founded Mensa, and received a knighthood. However, toward the end of his life a few raised doubts about his later data and articles. Shortly after his death many became convinced that his later work was fraudulent, with invented data to support his positions. Burt even wrote articles critical of Burt under pseudonyms, to which he then brilliantly replied (Hearnshaw 1979).

This raises the question of how to judge and trust experts. One form of the problem involves how the layperson can judge whether or not a self-proclaimed or even court-designated expert really is an expert. Another form of the problem involves the judgment by professionals and experts concerning other experts in other fields. Conflicts of professional turf and territory can bias the opinion of one sub-profession of another. Toxicologists or laboratory

microbiologists, for instance, may have differing opinions versus epidemiologists concerning a toxic substance or microbe.

There is also a problem concerning the relevant expertise of counter-experts of the sort that Beck and Giddens discuss. In the antinuclear movement and the opposition to biotechnology laboratories in some cities, some very eminent scientists lent their prestige and testimony to the opposition to corporations and government. However, sometimes the expertise of what the countermovement activists sometimes call “our experts” is not directly related to the specific issues involved. Linus Pauling, Nobelist expert in quantum chemistry, was the leading scientific spokesperson against nuclear testing, but was not himself privy to much of the secret data concerning the tests, while those, such as Edward Teller, who were privy, were defenders of the tests. For instance Mikio Kaku, a very able particle physicist and popularizer of science, and Barry Commoner, an organic chemist widely read and published on ecological issues, spoke against nuclear power. Both have some expertise relevant to the issues but are not nuclear engineers or students of low-level radiation. Similarly George Wald, a Nobel Prize winning biochemist lent his voice and prestige in opposition to genetic engineering labs, but was not himself a genetic engineer. (In this case several radical molecular biochemists involved in gene isolation, Jon Beckwith and Jonathon King also lent their voices.) Noam Chomsky, certainly the most eminent linguist has written against race difference IQ theories, but is not himself a psychometrician or statistician. David Layzer, a physicist, has training perhaps even further from the IQ debate. Richard Lewontin, who is expert on biostatistics also wrote extensively against racial IQ theories and studies, with more relevance. However, most working nuclear engineers support the value and relative safety of nuclear power, and most working psychologists of group differences in IQ believe in group differences in IQ (Snyderman and Rothman 1988). Otherwise they would not be working in that field.

Citizens attempt to judge the experts. This is often done on the basis of prestige, such as Nobel Prizes, even if not in the relevant field. Also inductive evidence is often reasonably used by the public in terms of the past track record of the relevant risk management experts. Philosopher of science, Wesley Salmon (1963) considers such inductive arguments valid forms of positive ad hominem or appeal to authority. However, if safety estimates were given in the past that later are admitted to be vastly lower than now accepted, citizens totally distrust the experts. Of course much smaller disagreements among experts or revisions of estimates are sufficient to make many members of the public total skeptics about science. Small errors or revisions have been successfully publicized by global warming skeptics, for instance. Likewise, insulting and aggressive statements about climate skeptics in private e-mails written by researchers, hacked, leaked and widely publicized have led many in Britain to doubt the whole, huge body of climate research.

One irony with respect to the special status of technical experts in technocracy is that several recent theories of expertise (Collins and Evans 2009; Collins 2010; Dreyfus 2008) base genuine expertise on tacit, non-explicit skills, involving language, embodiment, and social relationships, not on the explicit formal rules that are involved in technocratic rationality and technocratic expertise.

---

## Further Research

Research for the future concerning risk management in technocracy includes continuing surveys of public opinion concerning trust in technical on various issues, and analyses of the

components of trust and their weight. Empirical surveys or interviews of government and corporate officials to determine the amount of faith they have in the reports of their own technosphere would also be desirable, but more difficult to undertake. Perhaps slightly easier to undertake would be anonymous surveys or interviews with technical experts concerning their own opinion of how much acceptance the public, and institutional leaders, respectively have in their advice.

However, much of the further research on the issue of risk management in technocracy will consist not in empirical surveys and questionnaires, but in conceptual analysis of the nature of technocracy and the appropriateness of the term to contemporary societies and institutions. Certainly no nation is a complete or perfect technocracy. Technocratic and anti-technocratic tendencies exist to varying degrees in different nations, governmental departments, and branches of industry. We need to evaluate to what extent is the term technocracy relevant or applicable to contemporary societies and institutions, as a useful descriptive and analytical too, not as a kind of insult or accusations, as critical theorists and ecological activists sometimes use the term, or as the elder President George H.W. Bush used against his opponent Michael Dukakis during his presidential complain.

The main conceptual and empirical issues involve to what extent technocracy, not in the Platonic or Comtean sense of literal rule of experts, but in the weaker and more indirect sense of Galbraith's "technosphere." If technocracy is alive and well today at all, it is in this form.

To determine the extent to which the technosphere influences major societal decisions, one needs to estimate to what extent the alternative between which executives, generals, corporate offices, and other leaders choose are fully structured or provided by the technosphere. Using conceptual analysis, one might pose the counterfactual question: How would executives have decided if the input of the technosphere was not present? Of course, there are great difficulties with counterfactual history or social science, insofar as rigorous and accurate laws of the sort supporting counterfactuals in the natural sciences are generally not available. Some historians totally deny the usefulness or conceptual content of historical counterfactuals. Nevertheless, in recent years there has been a great deal of serious discussion of historical counterfactuals in the social sciences (Ferguson 2000; Lebow 2010). Sketching historical scenarios and tracing out plausible consequences of the absence of certain expert advice for executive decisions would be a valuable enterprise.

Another conceptual issue is the degree of technocratic versus participatory approach in various forms of risk deliberation. Deliberative democracy involves discussion of issues and programs. In deliberations concerning risk, experts have various possible roles. Sunstein (2002) would claim that experts should have a predominant role in risk deliberations, for Sunstein does not consider fears not concerned with death, disease, or injury (such as involuntary, unfamiliar nature of the risk) as irrelevant. Sunstein considers technocracy a central part of deliberative democracy. Others would consider citizen participation in discussion and evaluation of the expert opinions to be essential to genuine democracy. Radical critics of the present order see technocracy and democracy as totally opposed (Feenberg 1999). The dominant American form of public deliberation involves experts for the opposing parties (workers versus owners, neighborhood residents versus initiators of local technological project, environmental activists versus corporations or government). Many European forms of deliberation involve nonconfrontational, nonpublic discussion (Renn 2004). These different approaches have different possibilities for developing in the direction of technocratic dominance or democratic involvement. The confrontational, court-like testimonials, cross

examinations, and rebuttals of opposing experts can evolve into what was earlier dubbed a kind of pluralistic technocracy. The public's opinions can be sidelined. On the other hand, the nonpublic nature of the nonconfrontational discussion behind closed doors can lead to dominance by experts in the conversation and sidelining of the informally arrived at and nonpublic policy conclusions of the discussion.

Another largely conceptual but partly empirical issue is to what extent we evaluate various risk communication strategies as technocratic. Clearly the direct informational or warning approach is technocratic. However, there are degrees of understanding or empathy of the public fears concerning risk, whether justified or not. It is a matter of drawing the line between technocratic communication and genuine participatory communication.

Some questions for conceptual and empirical investigation are:

1. To what extent is technocratic risk management necessary in any industrial society?
2. To what extent is technocratic risk management replaceable by "democratic planning" of some sort not run by technocratic experts, but not totally uniformed about objective risks?
3. When technocrats incorporate risk perception research into their models, to what extent can they take perceived risks into account and acknowledge them without ceasing to be technocrats?
4. Similarly, when technocrats make risk communication a major part of their task, what are the parameters of simply propagandizing, persuading, or engaging in democratic deliberation?
5. To what extent have risk managers been able to recognize and acknowledge their own political and professional biases in their effect on risk estimates?

## References

---

- Ames B, Magaw R, Gold LS (1987) Ranking possible carcinogenic hazards. *Science* 236:271–280
- Aristoxenus, Pearson L (eds) (1990) *Elementa rhythica: the fragment of book II and the additional evidence for Aristoxenean rhythmic theory*. Clarendon, Oxford
- Bacon F ([1624] 1989) *The New Atlantis*. In: *The New Atlantis and The Great Instauration* (trans: Weinberg J), rev edn. Oxford Inc/Harlan Davidson, Oxford
- Bailes KE (1978) Technology and society under Lenin and Stalin: origins of the Soviet technical intelligentsia, 1917–1941. Princeton University Press, Princeton
- Beck U (1992) *The risk society. Towards a new modernity*. Sage, London
- Bell D (1960) *The end of ideology: on the exhaustion of political ideas in the fifties*. Free Press, Glencoe
- Bell D (1973) *The coming of post-industrial society: a venture in social forecasting*. Basic Books, New York
- Bird K, Sherwin MJ (2005) *American Prometheus: the triumph and tragedy of J Robert Oppenheimer*. Alfred A Knopf, New York
- Brzezinski Z (1970) *Between two ages: America's role in the Technetronic Era*. Viking, New York
- Carson R (1962) *Silent spring*. Houghton Mifflin, Boston
- Cassidy J (2010) *How markets fail*. Farrar, Straus, and Giroux, New York
- Collins H (2010) *Tacit knowledge*. University of Chicago Press, Chicago
- Collins H, Evans R (2009) *Rethinking expertise*. University of Chicago Press, Chicago
- Comte A ([1830] 1988) *Introduction to positive philosophy* (trans: Ferre F (ed)). Hackett, Indianapolis
- Connor S (2007) Hawking warns: we must recognise the catastrophic dangers of climate change. *The Independent*, 18 Jan 2007
- Cranor CF (2007) Toward a non-consequentialist risk analysis. In: Lewens T (ed) *Risk: philosophical perspectives*. Routledge, London, pp 36–53
- Crease R, Selinger E (2006) *Philosophy of expertise*. Columbia University Press, New York
- Dahrendorf R (1967) *Society and democracy in Germany*. Doubleday, Garden City

- Dijksterhuis EJ (1961) *The mechanization of the world picture*. Oxford University Press, New York
- Dilthey W, Makreel RA, Rodi F ([1883] 1989) *Introduction to the human sciences*. In: Makkreel RA, Rodi F (eds) Princeton University Press, Princeton
- Douglas M (1985) *Risk according to the social sciences*. Russell Sage, New York
- Douglas M, Wildavsky A (1982) *Risk and culture: an essay on the selection of technical and environmental dangers*. University of California Press, Berkeley
- Dowd D (1964) Thorstein Veblen. Washington Square Press, New York
- Dreyfus HL (2008) *On the internet*, 2nd edn. Routledge, London
- Duhem P ([1914] 1954) *The aim and structure of physical theory* (trans: Wiener PP). Princeton University Press, Princeton
- Elsner H (1967) *The technocrats: prophets of automation*. Syracuse University Press, Syracuse/New York
- Feenberg A (1999) *Questioning technology*. Routledge, London
- Findlay JN (1974) *Plato: the written and unwritten doctrines*. Humanities, New York
- Fuller S (1997) *Science*. University of Minnesota Press, Minneapolis
- Ferguson N (2000) *Virtual history: alternative and counterfactuals*. Basic Books, New York
- Gaiser K (1980) *Plato's enigmatic lecture on the good*. Prorusis 25:5–37
- Galbraith JK (1967) *The new industrial state*. New American Library, New York
- Gendron B (1977) *Technology and the human condition*. St. Martin's Press, New York
- Gigerenzer G (2000) *Adaptive thinking: rationality for the real world*. Oxford University Press, Oxford
- Gigerenzer G (2007) *Gut feelings: the intelligence of the unconscious*. Penguin, London
- Gigerenzer G (2008) *Rationality for mortals*. Oxford University Press, Oxford
- Glickman T, Gough M (eds) (1990) *Readings on risk. Resources for the Future*, Washington, DC
- Goodchild P (2004) *Edward Teller: the real Doctor Strangelove*. Harvard University Press, Cambridge
- Greider W (1987) *Secrets of the temple*. Simon and Schuster, New York
- Gross P, Levitt N (1994) *Higher superstition*. Johns Hopkins University Press, Baltimore
- Habermas J (1970) *Technology and science as ideology*. In: Habermas J (ed) *Toward a rational society*. Beacon, Boston, pp 82–122
- Habermas J (1973) *Theory and practice*. Beacon, Boston
- Harding S (1998) *Is science multicultural?* Indiana University Press, Bloomington
- Hayek FA (1955) *The counter-revolution in science: studies in the abuse of reason*. Free Press, Glencoe
- Hearnshaw LS (1979) Cyril Burt: psychologist. Hodder and Stoughton, London
- Heims SJ (1980) *John von Neumann and Norbert Wiener: from mathematics to the technologies of life and death*. MIT Press, Cambridge
- Herken G (2002) *Brotherhood of the bomb*. Henry Holt and Company, New York
- Horkheimer M (1972) *Critical theory* (trans: O'Connell MJ). Herder and Herder, New York
- Horkheimer M, Adorno T, Noerr GS, Jephcott E (eds) ([1948] (2002)) *Dialectic of enlightenment*. Stanford University Press, Stanford
- Jardine N (1973) *Francis Bacon and the art of discourse*. Cambridge University Press, New York
- Jasanoff S (1986) *Risk management and political culture: a comparative study of science in the policy context*. Russell Sage, New York
- Jasanoff S (2005) *Designs on nature: science and democracy in Europe and the United States*. Princeton University Press, Princeton
- Kahneman D, Tversky A (1973) *On the psychology of prediction*. Psychol Rev 80:237–251
- Kahneman D, Slovic P, Tversky A (1982) *Judgment under uncertainty: heuristics and biases*. Cambridge University Press, Cambridge
- Kant I ([1781] 1996) *The critique of pure reason* (trans: Pluhar WS). Hackett, Indianapolis
- Kaplan F (1983) *The wizards of Armageddon*. Simon and Schuster, New York
- Kowarski L (1971) *Scientists as magicians: since 1945*. Paper at Boston Colloquium for the Philosophy of Science, 26 Oct 1971
- Kramer HJ (1990) *Plato and the foundations of metaphysics: a work on the theory of the principles and unwritten doctrines of Plato*. SUNY, Albany
- Lapp R (1965) *The new priesthood: the scientific elite and the uses of power*. Harper and Row, New York
- Lash S, Szerszynski B, Wynne B (eds) (1996) *Risk, environment & modernity: towards a new ecology*. Sage, London
- Latour B (1987) *Science in action: how to follow scientists and engineers through society*. Harvard University Press, Cambridge, MA
- Layton ET (1971) *The revolt of the engineers: social responsibility and the American engineering profession*. Case Western Reserve University Press, Cleveland
- Lebow RN (2010) *Forbidden fruit: counterfactuals and international relations*. Princeton University Press, Princeton
- Leibniz GW ([c 1680] 1966) *Logical papers* (trans: Parkinson GHR). Oxford University Press, Oxford
- Leibniz GW ([c 1680] 2001) *Labyrinth of the continuum: writings on the continuum problem 1672–1686* (trans: Arthur RTW). Yale University Press, New Haven

- Leinweber D (2009) *Nerds on wall street: math, machines and wired markets*. Wiley, New York
- Leiss W (1972) *The domination of nature*. George Braziller, New York
- Leiss W (1994) Ulrich Beck risk society. *Can J Sociol* 19:544–547
- Leiss W (2001) In the chamber of risks. McGill-Queens University Press, Montreal
- Leiss W (2004) *Mad cows and mothers' milk*. McGill Queen's University Press, Montreal
- Leiss W, Chociolko C (1994) Risk and responsibility. McGill-Queens University Press, Montreal
- Levitt N (1999) *Prometheus Bedevilled*. Rutgers University Press, New Brunswick, NJ
- Lindsay RR, Shachter B (2007) *How I became a quant*. Wiley, Hoboken
- Lingua Franca (ed) (2000) *The Sokal hoax: the sham that shook the academy*. University of Nebraska Press, Lincoln
- Lowrance WW (1976) *Of acceptable risk: science and the determination of safety*. William Kaufman, Los Altos
- Lowrance WW (1986) *Modern science and human values*. Oxford University Press, Oxford
- Malinowski B (1954) *Magic, science, and religion and other essays*. Doubleday Anchor Books, New York
- Manuel FE (1962) *The prophets of Paris*. Harvard University Press, Cambridge
- Marcuse H (1964) *One-dimensional man*. Beacon, Boston
- Mayo D, Hollander R (1991) *Acceptable evidence: science and values in risk management*. Oxford University Press, New York
- Merton RK (1947) *Social theory and social structure*. Free Press, Glencoe
- Merton RK (1973) *The sociology of science*. University of Chicago Press, Chicago
- Mumford L (1967) *The myth of the machine: technics and human development*. Harcourt Brace Jovanovich, New York
- Murphy R (1994) *Rationality and nature*. Westview Press, Boulder
- Myrdal G (1960) *Beyond the welfare state: economic planning and its international implications*. Yale University Press, New Haven
- Patterson S (2010) *Quants*. Crown, New York
- Plato ([c 355 BCE] 1975) *Philebus* (trans: Gosling JCB). Clarendon Press, Oxford
- Plato ([c 380 BCE] 1992) *Republic* (trans: Grube GMA, Reeve CDC). Hackett, Indianapolis
- Polanyi M (1958) *Personal knowledge*. University of Chicago Press, Chicago
- Popper K (1945) *The open society and its enemies*. Routledge, London
- Popper K (1962) *Conjectures and refutations: the growth of scientific knowledge*. Basic Books, New York
- Poundstone W (1992) *Prisoner's dilemma*. Anchor Books, New York
- Quinton A (1980) *Francis Bacon*. Oxford University Press, New York
- Rechard RP (1999) Historical relationship between performance assessment for radioactive waste disposal and other types of risk assessment. *Risk Anal* 19:763–807
- Reisch GA (2005) How the cold war transformed philosophy of science
- Renn O (2004) The challenge of integrating deliberation and expertise: participation and discourse in risk management. In: McDaniels T, Small MJ (eds) *Risk analysis and society: an interdisciplinary characterization of the field*. Cambridge University Press, Cambridge, pp 289–366
- Rickert H ([1896–1902] 1986) *The limits of concept formation in natural science* (trans: Oakes G). Cambridge University Press, Cambridge
- Rickert H ([1899] 1962) *Science and history: a critique of positivist epistemology* (trans: Reisman G). Van Nostrand, Princeton
- Saint-Simon H (1952) *Selected writings* (trans: Markham F (ed)). Basil Blackwell, Oxford
- Salmon WC (1963) *Logic*. Prentice-Hall, Engelwood
- Shrader-Frechette KS (1985a) Risk analysis and scientific method. D. Reidel, Dordrecht
- Shrader-Frechette KS (1985b) Science policy, ethics, and economic methodology. D. Reidel, Dordrecht
- Shurkin JN (2008) *Broken genius: the rise and fall of William Shockley, creator of the electronic age*. Palgrave Macmillan, London
- Skidelsky R (2009) *Keynes: the return of the master*. Public Affairs, Washington, DC
- Slovic PB (2000) *The perception of risk*. Earthscan, London
- Slovic PB, Fischhoff B, Lichtenstein S (1981) Perceived risk, psychological factors and social implications. *Proc R Soc Lond A* 376:17–34
- Snyderman M, Rothman S (1988) *The IQ controversy, the media and public policy*. Transaction, Totowa
- Spinoza B ([1677] 2000) *Ethics* (trans: Parkinson GDR). Oxford University Press, Oxford
- Starr C, Whipple C (1980) Risks of risk decisions. *Science* 208:115–117
- Sunstein CR (2002) *Risk and reason*. Cambridge University Press, Cambridge
- Taleb NN (2004) *The black swan*. Random House, New York
- Taleb NN (2010) *Fooled by randomness*. Random House, New York
- Thaler RH, Sunstein CR (2008) *Nudge*. Yale University Press, New Haven
- Triana P (2009) *Lecturing birds on flying: can mathematical theories destroy the financial markets?* Wiley, New York

- Veblen T ([1921] (1983)) *The engineers and the price system*. Transaction, New Brunswick
- von Fritz K (1977) Pythagorean politics in southern Italy. Octagon Books, New York
- Weber M ([1904] 2001) *The protestant ethic and the spirit of capitalism* (trans: Parsons T) intro by Anthony Giddens. Routledge, London
- Weber M, Roth G, Wittich C (eds) ([1914] 1968) *Economy and society: an outline of interpretive sociology* (trans: Fischoff E). Bedminster Press, New York
- Weinberg A (1972) Social institutions and nuclear energy. *Science* 177:24–34
- Wilson B (ed) (1970) *Rationality*. Basil Blackwell, Oxford
- Wynne B (1982) Rationality and ritual: the windscale inquiry and nuclear decisions in Britain. British Society for the History of Science, Chalfont St. Giles



# Index

## A

Aarhus Convention, 767, 768  
Ability to Pay Principle, 900, 955–956  
Abortion, 741, 755  
ABS. *See* Antilock braking systems  
Absolute risk, 627, 628, 637–639, 652, 654, 655  
Absolute risk aversion, 118, 120, 122, 125, 128  
Absolute sense of safety, 61  
Abstract thought, 680, 687  
Abuse of research participants, 182  
Acceptability of risks, 881, 882, 886, 893, 896, 899, 905  
Acceptable research risk, 180, 181, 192, 193, 197, 199–205, 207  
Acceptable risk, 56, 57, 60, 61, 67, 71, 72, 74, 989–991  
Acceptance, 50–51, 781, 782, 784–785  
Accident analysis and prevention, 241, 244  
Accidents, 747, 749  
Accountability, 577, 580, 581, 586–588, 590–591, 595, 883, 884, 886, 892, 900–901  
Acknowledged inadequacies, 98, 99, 106  
Action guidance, 406, 422–423  
Acts, 391–394  
Actualism, 44–45  
Adding the pros and cons, 519  
Addition of utilities, 531, 532  
Additive aggregation functions, 529  
Additive independence, 531  
Additive representation, 529, 531, 532  
Additive utility representation, 529, 532  
Additive utility representation of preferences, 529  
Administered world, 1146  
Adverse effects, 966, 973  
Advertisements, 694, 695  
Aerosol, 323  
Affect, 93, 95, 96, 100, 102, 106, 678, 680–688, 822–824, 826, 827  
Affect heuristic, 679, 682, 683, 822, 824  
Affect infusion model (AIM), 707  
Affective cognition, 696, 697  
Affective cognitive reactions, 698  
Affective reactions, 698, 699  
Affect/reason continuum (A/R continuum), 709, 710  
Affect-reason-involvement (ARI) model, 709, 710  
Agency contracts, 594

Agenda 21, 768  
Aggregation, 48, 49, 916, 918  
– functions, 528–529, 536–537  
– of utilities, 520, 540  
Aggregation worry, 917–921, 926  
Agreement, 765  
Agreement principle, 186–187, 194–196, 198, 199  
AIM. *See* Affect infusion model  
AIR, 452, 457  
Alar, 1152, 1153  
Alcohol interlocks, 893  
Aleatoric, 89, 92–94, 97  
Allais, M., 566  
Allais' paradox, 116, 482–485, 505, 506, 508, 515, 526  
Ambient exposure laws, 808, 809  
Ambiguity, 349, 1080, 1081, 1084, 1088, 1094, 1106, 1107, 1109, 1111  
Ambiguity aversion, 581, 587, 590–591  
Ambiguous risk, 1080, 1087  
Ames, B., 1153  
Amygdala, 700, 701, 705, 706, 719, 823, 824  
Analytic, 679–684, 687, 688  
Analytical approach, 969  
Analytical methods of risk analysis, 801  
Analytic cognition, 698, 699, 702, 705, 709  
Anger, 679–682, 698, 702, 703, 706, 708, 709, 712–714  
Annual fatalities, 821  
Anonymous risks, 922  
Anscombe, E., 835  
Anscombe, F.J., 387  
Antagonism, 668, 672, 673  
Anthropology, 1097–1098  
Anticipatory emotions, 679, 697, 698, 707, 715  
Antilock braking systems (ABS), 251  
Antinomy, 505  
Antonym of risk, 60–63, 71  
A(H1N1) pandemic influenzae, 214, 216, 228, 230  
Apparent paradox, 505, 506, 515  
Appraisal, 679, 681–682, 699–701, 708  
Appraisal tendency framework, 708  
Appraisal-tendency theory, 681  
Approach-avoidance, 703, 705, 707, 708  
A priori probabilities, 330  
Archimedean axiom, 480  
Archimedean conditions, 420  
Archimedean property of real numbers, 420

- A/R continuum. *See* Affect/reason continuum
- Arenas for risk governance (ARGONA), 764, 765, 777–783
- Argumentation analysis, 30–32
- Arguments from precaution, 963, 965, 966
- ARI model. *See* Affect-reason-involvement model
- ARI Slice, 710
- ARI Solid, 710
- Aristotle, 28, 31, 835–837, 840, 843, 844, 846, 849, 850, 854, 855
- Arrhenius, S.A., 322
- Arrow–Debreu contingent securities, 127
- Asperger syndrome (AS), 705–707
- Aspiration level, 583
- Assessing regulations, 523
- Asymmetric information, 127, 129, 130
- Atmosphere, 322
- 9/11 Attack, 747
- Attitude, 662, 664–666, 669–673
- Attitude toward risk, 527
- Attitudinal scales, 730
- Attributes, 518–542
- Audit process, 221
- Aumann, R.J., 387
- Autonomy, 820–822, 824
- Availability effect, 746–749
- Available alternatives, 821
- Aversion to cycles, 542
- Aversion to risk, 118, 120, 122, 125, 128, 524–527, 532, 533
- Axiology, 424
- Axiom, 489
- of independence, 506
  - of probability, 479, 481
  - of rationality, 390, 396, 398
  - of structure, 390, 396, 398, 400
- B**
- Background risk, 861
- Backward induction, 513–515
- Backward-looking, 883–886, 892, 898, 900–901, 905
- Backward-looking responsibility, 883–886, 898, 905
- Bacon, F., 1140, 1144
- Balancing goods and risks, 910–912, 928
- Bank-run, 591
- Bankruptcy, 49–50
- Basal ganglia, 704, 705
- Base rate neglect, 646
- Basic components of risk, 201
- Basic interests, 190–194
- Basic intrinsic attitudes, 532, 533
- Basic intrinsic aversion, 532, 533
- Basic intrinsic desire, 532, 533
- Basic rights, 955, 956
- Bayesian, 65–66, 72, 148, 150, 152, 153, 157, 447, 454, 468, 473
- approaches, 330
  - conditionalization, 399
  - decision theorists, 424
  - decision theory, 376–401, 415, 424, 509, 518, 519
  - Event Trees, 360
  - learning, 584
  - methods, 351, 361
  - revision, 37
- Bayesianism
- objective, 399
- Bayes' rule, 139–142, 148
- Bayes' theorem, 140, 399, 424, 639, 640
- Beck, U., 1004, 1009, 1012–1014, 1022
- Behavioral assumptions, 228–230
- Behavioral axiom, 480
- Believe-based view of trust, 863
- Bell, D., 1139, 1143
- Belmont Report, 181, 183
- Benefits, 47–49, 911–913, 918, 921, 923–927, 929
- Bernoulli, D., 378
- Bernoulli's hypothesis, 408–410, 428–429
- Betrayal aversion, 593–594
- Bets, 388–390, 392–393, 398, 479–482, 485
- Betting method, 479, 480
- Bias(es), 154–156, 509, 821, 823, 825–828
- Biased assimilation and polarization, 739, 742–746, 752, 753
- Biased reporting, 623, 627–629, 652, 654
- Binary measures, 66
- Biomedical research, 179–208
- Biomonitoring, 811
- Biotechnology, 34
- Bisphenol A, 812
- Biosphere, 322
- Blameworthiness, 883–886, 900–901, 905
- Bolker, E.D., 394
- Borch's condition, 119
- Borderline personality disorder (BPD), 706, 707
- Boundaries, 1100, 1101, 1110, 1111
- Boundary work, 1111, 1124, 1127–1128, 1133
- Bovine Spongiform Encephalopathy (BSE), 1148, 1157
- BP, 36
- BPD. *See* Borderline personality disorder
- Brain implants, 825
- Branding, 694, 695, 716
- Breathalyzer, 244–245
- British empiricism, 1140, 1144
- British empiricists, 1144
- Broca's area, 702, 703
- Brokerage, 587
- Broome, J., 910, 913–915, 926
- Brzezinski, Z., 1143

- BSE. *See* Bovine Spongiform Encephalopathy  
BSE crisis, 1105  
BSE (or mad-cow disease) crisis, 1120, 1121, 1130  
Buckyballs, 490–492  
Burden of proof, 38, 968, 971, 972  
Bush, George H.W., 1159
- C**
- Cancer, 97–99, 102–104, 107  
Capability approach, 980–996  
Capacity-responsibility, 882, 883  
Capitalism, 50  
Carbon capture and sequestration (CCS), 105–109  
Carbon capture and storage (CCS), 100, 101, 105–107, 820, 828  
Carbon dioxide, 320–323, 332  
Carcinogen, 808, 810, 811, 813, 815  
Cardinal scale, 413, 428  
Cardinal utility, 116  
Cardinal utility scales, 413  
Carson, R., 1149  
CASC scale. *See* Communication via analytic and syncretic cognition scale  
Catastrophic vs. chronic risks, 821  
Catch-all probability, 445  
Causal decision theory, 401  
Causal dilution, 44–47  
Causal-responsibility, 882, 897, 898  
Causation, 34  
Cautious shift, 580  
CBA. *See* Cost-benefit analysis  
CCS. *See* Carbon capture and sequestration; Carbon capture and storage  
CDR. *See* Common, but differentiated responsibilities  
Centipede game, 513  
Cerebral lateralization, 702, 707  
Certainty, 410, 413, 414, 417, 418, 425, 427, 428, 430, 437  
Certainty effect, 483  
Certain uncertainties, 1107  
Character, 835–845, 850, 851, 854  
Childhood contamination, 806, 816  
Choices of general model, 106  
Choices under certainty, 410  
Chomsky, N., 1153, 1158  
Cingulate gyrus, 704, 705  
Classical approach, 1071–1072, 1076, 1079, 1082  
– to TA and RA, 1071, 1075  
– to TA and risk, 1070, 1075, 1081, 1090  
Classical interpretation of risk, 1074–1079  
Classical risk approach, 1070–1077, 1079–1081, 1084, 1085  
Classical TA, 1072–1075, 1077–1083, 1085  
Classical (parliamentary) TA, 1074  
Classification, 88–93, 96, 97, 108, 109
- Climate change, 320–337, 594, 596, 736, 737, 741, 742, 747–749, 753–755, 802, 827–829, 897, 900–901, 933–935, 937, 938, 940, 942, 945–947, 949–951, 954–956  
Climate models, 320–325, 328–330, 332  
Climate projections, 323  
Climate risks, 892  
Clinical equipoise, 183–184, 187, 189, 192, 199  
Clinical research studies, 180  
Clinical trials, 49  
Cloning, 820, 825  
Closed fuel cycle, 298, 303, 308, 311  
Closest deterministic analogs, 43  
Codes of ethics, 887–888  
Coefficient of variation, 527, 528  
CO<sub>2</sub>-emissions, 827  
Cognition, 824  
Cognition-based affect-based trust, 872  
Cognition-based trust, 872  
Cognitive biases, 186, 202  
Cognitive decision, 546, 552  
Cognitive development, 706  
Cognitive effort, 587  
Cognitive emotions, 709, 710  
Cognitive processing, 699–707, 709  
Coherent preferences, 529–531  
Coherent set of beliefs, 481  
Cold processing, 710  
Collective actions, 794  
Collective agency, 892, 897–900, 902  
Collective decisions, 894, 895, 899  
Collective responsibility, 171, 172, 887, 890–896, 898, 905  
Collective utility, 536, 537  
Columbine school-shooting massacre, 747  
Column disutility, 567  
Combining attributes' utilities, 528  
Combining evidence, 150–152  
Common belief, 514  
Common, but differentiated responsibilities (CDR), 900  
Commoner, B., 1149, 1158  
Common knowledge, 511, 513–514  
Communication, 762–785, 823, 825, 828, 1094, 1095, 1098, 1103, 1104, 1107–1112  
Communication via analytic and syncretic cognition (CASC) scale, 709, 712  
Communism, 1142  
Communitarian, 727, 732, 733, 740, 741, 743, 749, 751, 752, 756  
Comparative decision theory, 552, 557–559, 561, 563, 564  
Comparator activity, 191, 192, 202, 207  
Comparison, 917, 919, 920, 925, 927  
Comparisons of options, 520–522, 540  
Compassion, 820, 824, 829

- Compensation, 792–796, 798, 800, 920, 923, 927–929, 951, 956  
 Competence, 577  
 Competent and informed decision-maker, 194–196  
 Complementarity, 536  
 Complete, 411, 412, 415, 417, 422, 425, 426, 428, 502  
 Completeness, 411, 412, 422  
 Complex, 1069, 1070, 1075, 1076, 1080, 1081, 1087  
 Complexity, 584, 587, 588, 592, 594–597, 1073, 1075, 1076, 1080, 1081, 1094–1097, 1099, 1101–1103, 1106–1122  
   – layers of, 763  
   – reduction of, 865  
 Complex systems, 321, 326–331, 335–336  
 Component analysis, 186–193, 196, 198–200  
 Compositionally and dynamically complex systems, 327  
 Compound lottery, 385, 387  
 Comprehensive doctrines, 903  
 Comprehensive utilities, 520–522, 529, 530, 533, 536, 537, 540  
 Comte, A., 1140–1141, 1143, 1151  
 Concatenation operation, 533  
 Conceptions of liberty, 893  
 Conceptions of risk, 879–886, 889, 905  
 Concept of responsibility, 882, 885, 901–903  
 Conceptualization uncertainty, 321  
 Concerns, 820–822, 825–829  
 Concerns for “third parties”, 775  
 Conditional expectation value, 394, 395  
 Conditional independence, 147  
 Conditional probabilities, 330, 336  
 Condition for acceptable research, 204, 205  
 Condorcet, 51  
 Confidence, 858, 860, 861, 865, 867, 868, 871  
 Confidence interval, 152  
 Conflict, 669  
 Conflicts of interest, 827  
 Conformity, 577, 580, 582, 584, 585  
 Consent, 46–48, 50, 51, 181–183, 186, 188, 189, 192, 194–207, 792, 794–800, 920, 923–925, 927–929  
 Consequences, 980, 982, 984–987, 990, 991, 994, 995  
 Consequentialism, 425, 835, 839, 1144  
 Consequentialist, 838, 839, 843, 845–846, 855  
 Consequentialist ethics, 1144  
 Constant acts, 392  
 Constrained, 190, 193, 198, 205  
 Constructive process, 686  
 Constructivism, 1004, 1016, 1020–1022  
 Constructivist, 1100  
 Consumer preferences, 799  
 Consumer Product Safety Act, 806  
 Consumption smoothing, 122  
 Contamination, 806, 811–816  
 Contents and requirements, 782  
 Context effect, 154–157  
 Contingent valuation method, 799  
 Continuity, 411, 420  
 Continuity axiom, 385  
 Continuous data, 147, 152  
 Contract theories, 46  
 Contractualism, 917–918  
 Control, 34–36, 40, 43, 49, 51, 61, 66–68, 82, 250, 252–255, 257, 258, 260, 878, 881, 885, 886, 897–899, 902, 905  
 Controlling risks, 861  
 Control mechanisms, 779, 782  
 Controversial values, 30, 32  
 Convex set, 550  
 Cooperating, 512  
 Cooperative behavior, 863, 868, 869, 871–873  
 Cooperative breeding, 873  
 Coordination, 590–592, 596  
 Coordination games, 591–592  
 Corpus callosum, 702  
 Cosmetic ingredients, 811–813  
 Cost-benefit analysis (CBA), 60, 821–824, 933, 946–953, 971, 989, 991, 992  
 Council of Newton, 1141  
 Creativity, 488, 496  
 Credibility, 34, 739, 749–752  
 Credible interval, 152, 153  
 Critical theory, 1145, 1146, 1148–1150  
 Cross-cultural risk perception, 737–738  
 Cross-disciplinary, 781, 782  
 Cryosphere, 322  
 Cultural, 93, 96  
   – approach, 56–58, 70, 78–79  
   – availability, 739, 746–749  
   – biases, 881, 885  
   – cognition, 725–757  
   – credibility, 739, 749–752  
   – identity affirmation, 739, 752–753  
   – norms, 680  
   – theory, 254, 662, 664, 665, 672, 881, 1097  
   – worldviews, 726, 729, 733, 735–737, 739–746, 749, 752, 753, 755  
 Cultural Cognition Project, 726, 752–755  
 Cultural theory of risk, 725–757  
 Culture, 1031, 1032, 1041, 1042  
 Cumulative development of knowledge, 771  
 Cumulative net risks, 198  
 Curiosity, 711  
 Currency, 680  
 Cyborgs, 160, 161, 163–164, 174–175, 825  
 Cyclical choices, 542  
 Cyclical preferences, 422

**D**

- Dahrendorf, R., 1143  
 Danger, 1010–1012, 1015–1016, 1023  
 Davy lamp, 256

- Davy, Sir Humphry, 256  
Decision
  - under certainty, 59, 415, 417
  - under risk, 59, 396, 400, 415, 547
  - under uncertainty, 59, 415, 422, 429, 437
  - making, 478–483, 485, 487, 489, 490, 493–497, 695–699, 701, 702, 706–708, 711–712, 715
  - matrix, 416
  - process, 766–768, 775–776, 779, 780, 783–785
    - steps, 784
    - rules, 552–556, 559, 563, 564, 569, 571
    - shared, 625, 629, 650, 654
  - Socratic approach, 478
  - theory, 29, 35–36, 39, 43–45, 47, 334, 880
- Declaration of Helsinki, 182, 183  
*De Dicto Worse*, 941, 942  
Deep brain stimulation, 491  
Deep ecology, 326  
Deepwater Horizon, 1155  
Deepwater Horizon oil rig, 878  
Defeasance problem, 47–48  
Defense attorney's fallacy, 143  
Defensive decision making, 625, 627, 652  
Deficit model, 1007, 1017–1019  
The Deficit model of public understanding of science, 1007  
De Finetti, B., 388, 391  
Definition, 29–31, 38, 43, 49
  - of risk, 58–59, 62
  - of trust, 860
- Degree of structure, 770–771  
Degrees of belief, 381, 382, 387–390, 393, 399, 479, 481, 485  
Degrees of desire, 520, 531–533  
Deliberation, 890, 1070, 1083, 1085–1086, 1088, 1089  
Deliberative processes, 799–800  
Delimitation issues, 69  
Democracy, 50–52, 767, 772, 778, 780, 784–785, 822, 1004, 1013, 1016, 1017, 1019–1020, 1024  
Denial, 699–700  
Deontology, 835, 843, 855
  - respect-for-persons judgments, 823
  - theories, 46
- De Re Worse*, 941  
de Saint Simon, H., 1140, 1141  
Descartes, R., 1140, 1144  
Descriptive, 1102  
Descriptive decision theory, 540  
de-Shalit, A., 981, 985, 993, 995  
Design, 37, 39, 41–43  
Desire-neutral, 389, 391  
Desires, 376, 380–384, 388–392, 394, 401  
Determinism, 92, 96  
Deterministic, 93  
Deterministic (approach to risk prevention), 276  
Determinists, 93, 94  
Deus é Brasileiro, 260  
Developing children, 811–814, 816  
Development, 983, 986, 989, 993–994  
Developmental interactionist (DI) theory of emotion, 715  
Developmental origins of health and disease, 811  
Developmental toxicants, 811–813  
Diagnostic, 687, 688  
Dialectic, 1139, 1145–1146, 1148, 1156  
Dialectical reason, 1145, 1146, 1148  
Dialogues, 762, 763, 765, 766, 768–769, 771, 772, 782, 828  
Diamond's case, 430  
Dianoetic, 337  
Dichotomy between reason and emotion, 820, 823, 828  
Dictator game, 586, 592  
Differences, 771, 773, 776–779  
Different number choice, 938–942, 944  
Dilthey, W., 1147  
Dimension of attributes, 522  
Dimension of possible outcomes, 522  
Diminishing marginal utility, 525, 526  
Diminishing sensitivity, 580  
Disappointment, 679  
Disaster response chain, 343  
Discounting, 946, 949–951  
Discount rates, 802, 933, 949–952  
Discursive dilemmas, 898  
Discursive psychological, 1053, 1054  
Disease transmission, 218, 224  
Disgust, 703, 704, 709, 825  
Display rules (pseudospontaneous communication), 704  
Dispositions, 730, 731, 735, 738, 741  
Dissent, 1069, 1080, 1081  
Distal, 664  
Distinctive, 921, 923, 927, 928  
Distinctive between therapeutic and non-therapeutic research interventions, 182, 183, 188, 189, 193, 199  
Distracted driving, 240, 241  
Distribution, 800, 801, 910–929
  - fair social system, 925
  - impartial, 913–914, 916
  - indiscriminate, 915, 916, 922
  - unequal, 911, 913, 921–928
- Distribution of goods, 910–913  
Distribution of harm vs. probability of harm, 912, 913, 916  
Distribution of responsibility, 893, 900, 903  
Distribution of risk and benefit, 796, 800–801  
Distribution of risks, 911–913, 918  
Distributive aspects, 68, 73, 76  
Distributive justice, 286–291, 792, 800  
DNA profiles, 141–144, 149–151, 155  
Doctrine of the double effect, 925  
Domain-irrelevant context information, 138

- Domestic terrorist, 747  
 Dominance, 510–512  
 Domino effects, 775  
 Donations, 586  
 Dose-effect relations, 889–890  
 Double blind, 586  
 Douglas, M., 254, 1004, 1010–1012, 1150, 1151, 1153, 1155  
 Dread, 678, 682, 685, 822, 825  
 Dredging company, 860  
 Drinking and driving, 244, 245  
 Driving furiously, 241  
 Dual-process theories, 678–679, 681, 682, 783, 823  
 Dual use, 454, 455  
 Dual use dilemma, 170–171  
 Dukakis, M., 1159  
 Dunning–Kruger effect, 613–614  
 Dutch Book, 422, 503  
 Dutch Book-arguments, 398, 399  
 Dutch Book theorem, 479, 481  
 Dynamic consistency, 582, 583
- E**
- Each person into, 917  
 E-admissibility, 486–488  
 Earthquake recurrence, 359  
 Ecological validity, 688  
 Economic approaches, 933, 946, 951, 952  
 Economic theory, 48  
 Economic value of human life?, 798  
 Edgeworth, F.Y., 1144  
 Education, 741, 742, 745, 748  
 EFSA. *See* European Food Safety Authority  
 Egalitarian, 254, 255, 727–730, 732, 733, 736, 740, 741, 743, 749, 751–752, 756  
 Egoistic emotions, 829  
 EIA directive. *See* Environmental impact assessment directive  
 Electric Power Research Institute, 1153  
 Electro-magnetic fields, 490, 494  
 Ellsberg's paradox, 482–485, 487–489, 508, 509, 566, 581  
 “Embodied” experiences, 681  
 Embryonic stem cells (ePS-cells), 485  
 Emerging infections, 215, 216  
 Emolumentum, 409  
 Emotional, 820–830
  - appeals, 694–699, 708, 711
  - communication, 703, 706
  - competence, 715–717
  - education, 715–717
  - experiences, 697, 698, 701, 703, 707, 715
  - involvement, 709
  - message, 712
  - risk communication, 696
 Emotional-education approach, 716  
 Emotional influences, 712–715
- Emotion-focused, 700  
 Emotion relative to reason, 714  
 Emotions, 665–667, 678–683, 687, 701–702, 708, 709, 711, 712, 714, 716, 717, 844–847, 849, 852–853, 855
  - incidental to the decision, 696–697
  - intrinsic to the decision, 696–697
 Empathy, 706, 707, 824–826  
 Empirical decision theory, 821  
 Empirical judgment, 184  
 Empirical philosophy, 873  
 Encapsulated interest view, 864  
 Encounters with risk, 216, 218, 220, 221, 231–233  
 Energy crisis speed limits, 243, 245  
 Energy scenarios, 323, 324  
 Engels, F., 1142  
 Engineering ethics, 887  
 English sayings, 964  
 Enlightened engineering and psychology, 1096–1097  
 Environment, 827  
 Environmental Impact Assessment (EIA) Directive, 768, 778  
 Environmental law and policy, 962, 966, 967  
 Environmental policies, 801  
 Environmental Protection Agency (EPA), 306–307, 315  
 Environmental risk, 728, 735, 740–742, 746  
 EP. *See* Exceedance probability  
 EPA. *See* Environmental Protection Agency  
 Epidemic outbreak, 218, 232  
 Epidemiological, 214–216, 220, 224–230
  - mechanisms, 225, 230
  - models, 216, 228
 Epigenetic effects, 813  
 Epistemic, 89, 92–94, 97
  - certainty, 492
  - reliability, 551
  - risk, 478–497
  - rule, 971–973
  - state, 492–494, 496
  - trust, 668, 673
  - trust and social trust, 780
  - uncertainty, 56, 66–69, 72, 73, 76, 80–82, 478, 487, 490–497
  - values, 32–33
  - virtues, 337
 Epistemological, 93  
 Epistemology, 34–35  
 ePS-cells. *See* Embryonic stem cells  
 EQ. *See* Equality claim  
 EQCAT, 452  
 Equal distribution, 910–917, 920, 921, 923, 928  
 Equality, 192–194, 205, 910–914, 917, 921, 928, 929  
 Equality claim (EQ), 913–915, 917  
 Equality concerns, 582, 584  
 Equal opportunity, 297, 302, 304, 310, 313  
 Equal treatment, 917, 920, 928

- Equations, 321, 326  
 Equilibrium selection, 591  
 Equiprobability assumption, 330–331  
 Equitable social system, 925  
 Equivalent attributes, 529  
 Equivocation, 563, 565  
 Error of type I and type II, 32, 33  
*Esperance Morale* (moral expectation), 409  
 Estimative system, 699  
 Estrogen-mimicking, 811, 812  
 Ethical acceptability of risks, 899  
 Ethical analysis, 1077, 1078  
 Ethical and social impacts, 1070, 1076  
 Ethical issues, 168, 176  
 Ethically neutral, 800, 803  
 Ethically neutral proposition, 479–480, 482  
 Ethics, 29, 33, 43–49, 52  
   – of risk, 881, 895  
 Ethno-science, 1155  
*Eudaimonia*, 836, 837  
 Eudoxus, 1139  
 EU risk regulation, 1119–1134  
 European Commission, 1121, 1122, 1124, 1129, 1131  
 European Food Safety Authority (EFSA), 1122,  
   1124–1131  
 European Parliament, 1121, 1122  
 European revolutions of 1789 and 1848, 324  
 Evaluative decision theory, 540  
 Event, 416, 417, 419, 420  
 Evidence lineup, 156  
 Evidential, 510  
   – decision theory, 401  
   – probability, 526  
   – risks, 526, 527  
 Evolutionary theory, 707, 708  
 Evolution of cooperation, 870  
 Ex ante, 184  
 Exceedance probability (EP), 457, 464, 466  
 Excessive, 180, 181, 184, 186, 189, 190, 193, 195–200, 203,  
   206, 207  
 Exemplar narratives, 504  
 Expectational utility function, 529  
 Expectation transitivity, 564, 565  
 Expectation value, 30, 58–60, 64, 69, 71  
 Expected monetary value, 377–379, 397  
 Expected utility, 30, 36, 43, 45, 114–116, 128, 129, 376,  
   379, 381–390, 393, 395–401, 478–480, 483, 484, 486,  
   487, 489, 493, 518–520, 522, 525, 529, 533–535, 678  
   – principle, 520, 533, 534  
   – theorem, 116  
   – theory, 579, 580  
 Expected value, 483, 487, 488  
 Experienced emotions, 697, 698, 701, 703, 707, 715  
 Experiential, 679–683, 685, 688  
 Experimental animals, 808, 812, 815  
 Experimental animal studies, 808, 812  
 Expert assessments of risk, 534  
 Expert-citizen/consumer interactions, 1050, 1053  
 Expertise, 35  
 Expert judgement, 358, 359, 365  
 Experts, 137–139, 141–146, 149, 150, 152–156, 662, 663,  
   669–673, 747–752, 754, 820, 821, 824, 826–828, 830,  
   990, 991, 1121, 1123–1127, 1129–1133  
 Explanation-based predictions, 216–217, 221–226,  
   230–233  
 Explanatory power, 32, 664, 665, 670  
 Exposure, 860, 867  
 Exposure assessment, 808, 814, 815  
 External normativity, 71  
 Extortion, 795  
 Extrinsic desire, 532
- F**
- Facial displays, 704  
 Fact of irrationality, 494, 495  
 Factor transitivity, 564  
 Facts box, 625, 626, 653  
 Fact–value distinction, 30, 32–34, 38, 39  
 Fair distribution, 821, 822, 910–929  
 Fairness, 577, 581–583, 588, 910–914, 917, 920–929  
   – of a die, 914  
   – vs. risk reduction, 912, 913, 917, 921  
 Fair social system, 925  
 Fallacies, 31–32  
 False negative, 32, 889  
 False-positive, 32, 889  
   – test results, 629, 630  
 Falsidical paradox, 505, 510, 511, 515  
 Fatalists, 253–255, 729, 735–736  
 Fear, 678, 680–682, 694, 698, 700, 702, 703, 705, 708, 709,  
   713, 822, 825, 829  
 Fearful individualistic emotions, 713  
 Fear of flying, 825  
 Fear or anxiety, 794  
 Feedbacks, 321–323, 326, 329  
 Feelings, 677–688, 822, 824, 825, 827, 829  
   – of others, 716  
   – own, 716  
   – of responsibility, 827, 829  
 Feelings-based and cognitive processes, 681–682  
 Fiction, 824, 829  
 Fight fires, 926–927  
 Final model outcome, 96, 97  
 Financial agency, 594  
 Financial incentives, 587  
 Financial payment, 206  
 Financial risks, 829  
 Financial sector, 829  
 Fingerprint, 138, 144, 145, 151, 155, 157  
 Finitude, 552

- Fire fighters, 920, 926  
 Fission products, 298, 309  
 Flash crash, 1152  
 Floor of the ARI solid, 710  
 Food and Drug Administration (FDA), 807  
 Forensic expert, 137–138, 141, 144, 145, 149, 151–154, 156  
 Forensic statistics, 138, 153  
 Formal methods, 973–974  
 Forms of participation, 769  
 Forward-looking moral responsibility, 879  
 Forward-looking responsibility, 883–886, 892, 898, 900, 902, 905  
 Foucault, M., 214, 215, 219, 232, 234, 1023  
 Framed, 894, 896  
 Frames, 1019, 1020, 1024  
 Framing effects, 796, 799  
 Framings, 581, 583, 584, 586, 596, 753, 755, 1009, 1018–1020, 1024  
 Frankfurt school, 1146  
 Freedom of choice, 1059  
 Frequencies, 324, 330  
 Frequentist, 153  
 Friends of the Earth, 1126, 1128  
 FTT. *See* Fuzzy-trace theory  
 Fukushima, 827  
 Fullerenes, 490, 491  
 Fuller, S., 1149–1150  
 Fully comparable groups, 564  
 Functionalist, 739  
 Functional magnetic resonance imaging, 685  
 Functioning, 981–984, 986–989, 993, 994, 996  
 Fundamentally justify, 204–206  
 Future generations, 800, 802, 932–937, 940, 941, 946, 949–950, 953, 954, 957  
 Future people, 932–934, 936, 938–946, 949–951, 955–957  
 Fuzzy-trace theory (FTT), 682
- G**  
 Galbraith, J.K., 1143, 1144  
 Gambling devices, 384, 387, 389, 395  
 Games of perfect information, 514  
 Game theory, 413–415, 481, 511, 512, 514, 538  
 Gap between experts and public, 820  
 Gardoni, P., 983, 985–990, 992–995  
 Gender, 741, 742, 745, 1029–1047  
 General circulation models (GCM), 323  
 General framework for risk-benefit evaluations, 181, 194, 204  
 General idea of precaution, 963–965, 968  
 Generalized decision principles, 540  
 Generalized decision theory, 486, 489  
 Generalized expected utility, 554  
 Generalized “normative” decision theories, 485
- General suspiciousness, 672  
 Genetically modified food, 665, 666, 671  
 Genetically modified organisms (GMOs), 490, 492, 494, 1124–1131  
 Genetic therapy, 523  
 Genetic variation, 814  
 Genuine uncertainty, 802, 803  
 Geological disposal, 297, 299, 301–302, 304, 306–314, 316  
 German Green Party, 827  
 Gigerenzer, G., 1154  
 Globalization, 342  
 GM-foods, 820  
 GMOs. *See* Genetically modified organisms  
 Goal-directed behavior, 701  
 Goodwill, 866, 867  
 Governance, 214, 215, 218, 219, 221, 223, 231, 233–235, 768, 777, 778, 780, 782, 783, 1093–1112  
 Governmentality, 214, 215, 231, 233–235, 1002, 1004–1005, 1016, 1022–1023  
 Government regulatory decisions, 536  
 Graph literacy, 631, 645, 647, 650  
 Great depression, 1142  
 Greenhouse effect, 34, 38  
 Greenhouse gas, 320–323, 325, 326, 331, 332, 334, 337  
 Greenpeace, 1125, 1128  
 Grey goo, 160–162, 173–174  
 Grid-group typology, 1011, 1012  
 Gross, P., 1152  
 Group and grid, 727–731, 733–735, 737–739, 741, 750, 756  
 Group discussions, 765, 768–771, 773, 775  
 Group polarization, 739  
 Guardian, 765  
 Guilt, 824  
 Gun, 736, 738, 741–742, 745, 747, 749, 756  
 Gut feelings, 1154  
 Gut reactions, 823, 828
- H**  
 Habermas, J., 1146–1149  
 Haemophilus influenzae type b (Hib) bacteria, 216–217, 221–226, 229, 232  
 Haight, F., 241  
 Half-life of a substance, 811–812  
 Hansson, S.O., 913, 924–926  
 Happiness, 501  
 Hard impacts, 1052, 1058–1065  
 Harm, 56, 59–76, 80–82, 792–798, 800–801  
 Harmful condition, 933, 941–943, 946  
 Harm *vs.* probability of harm, comparison, 913  
 Harsanyi’s theorem, 529–531  
 Hawking, S., 1157  
 Hazard, 808, 880, 881, 888, 898, 900  
 Hazardous facility siting, 286

- Health, 827  
   – literacy, 627, 630, 633  
   – numeracy, 631  
   – risks of radiation from mobile phones, 895
- Hedge fund, 1143, 1151, 1152
- Hegel, G.W., 1145, 1146
- Hematopoietic stem cell transplantation, 493
- Hempel, C., 32
- Hermeneutics, 1147
- Heuristics, 749–752, 756, 821–824
- Heuristic thinking, 698
- Hi and Lo, 862
- Hidden value assumptions, 29, 32
- Hierarchical, 728, 733, 735, 741, 742, 746, 749, 751, 752, 755
- Hierarchists, 254, 255
- Hierarchy-egalitarianism, 729–732, 736
- Hierarchy of propositions, 149
- Higher order probabilities, 479, 486, 488
- Higher order uncertainty, 490
- High-impact, low-frequency events, 226
- High level waste (HLW), 298, 299, 309–311
- High road, 701
- Hindsight bias, 36
- HLW. *See* High level waste
- Hobbes, T., 792, 858, 859
- Holistic-syncretic, 699, 702
- Homo economicus, 490
- Homogeneous groups, 564, 565
- Hope, 678
- Horkheimer, M., 1146, 1149
- Horse-race lotteries, 387
- Hot processing, 709–710
- Household decision making, 589
- How safe is safe enough?, 797, 798
- HPV. *See* Human papillomavirus
- HPV vaccine, 750, 751
- Human-animal hybrids, 825
- Human enhancement, 163, 174
- Human error, 41
- Human papilloma virus (HPV), 627–629, 750
- Human rights, 950, 955, 956
- Hydrosphere, 322
- Hypotheses, 136, 138–142, 144–154, 156
- Hypothetical, or implicit justification, 797
- I**
- IAEA. *See* International Atomic and Energy Agency
- ICD. *See* Internal cardiac defibrillators
- Icon arrays (pictographs), 645–647, 653
- Idealizations, 519, 537, 538, 547, 548, 550, 553
- Ideal rationality, 548
- Ideal social arbiter, 200
- Identity, 1053, 1055–1056, 1058, 1061, 1062
- Identity-and lifestyle related concerns, 1055–1056
- Identity-protective cognition, 739–743, 746, 752
- Ideology, 739, 741, 742, 745
- Idiographic, 1147
- “If we don’t someone else will” argument, 173–174
- Ignorance, 88, 90–93, 96, 349, 407, 415, 1077
- IIHS 2007, 251
- Illusion of certainty, 625, 630, 639, 647–649, 654
- Imagery, 683, 687
- Images and feelings, 679, 683
- Imitation, 594, 595
- Immediate emotions, 679, 696
- Impact models, 321, 324–326, 328, 332, 333, 335–336
- Impartial, 916
- Impartial distribution, 914
- Impartiality, 913–915
- Imperative of responsibility, 333
- Imperfect knowledge, 272, 279
- Implementation, 964, 970–973, 975
- Implicit assumptions, 94, 98, 104
- Implicit consent, 798
- Imposed risks, 252, 253, 256
- Imprecision in probability and utility assignments, 538
- Incentive values, 701
- Incidental, 679, 680, 686–688
- Incidental emotions, 679, 696–697
- Inclusion, 1095, 1105, 1108–1112
- Incoherence, 972, 973  
   – choices, 542  
   – sequences of choices, 541, 542  
   – sequences of decisions, 538
- Income, 741, 742, 745
- Incommensurability of the attributes, 539
- Incomparable attributes, 531, 536, 539–542
- Incomparable options, 531, 539–542
- Incomplete orderings of options, 539
- Indecision, 561
- Independence, 529, 531, 536, 552, 558, 563, 565–571  
   – assumptions, 481, 483, 484  
   – axiom, 385, 392, 481, 582  
   – condition, 502, 507
- Indeterminacy, 88, 91, 98, 349
- Indeterminacy of preferences, 539
- Indicators, 983, 987, 993–995
- Indifference, 384, 389, 390, 561
- Indifference principle, 511
- Indignation, 825
- Indiscriminate risk distribution, 915–917
- Indiscriminate risk impositions, 915–917
- Individualism-communitarianism, 730–732
- Individualist emotions, 711
- Individualistic, 728–730, 733, 735, 736, 741–743, 745, 746, 749, 751–753, 755, 756
- Individualistic emotions, angry, 713

- Individualists, 249, 250, 252, 254–257, 260  
 Individualization, 1004, 1013, 1014, 1022  
 Individual liberty, 195  
 Individual responsibility, 890–893, 895, 896  
 Individual rights, 923–924  
 Individual's perspective, 800  
*In dubio pro natura*, 962  
 Induced pluripotent stem cells (iPS-cells), 485  
 Inductive method, 1140  
 Industrial chemicals, 806, 810–816  
 Industrialism, 1141  
 Inequality, precautionary measures, 929  
 Infectious risk, 216, 217, 221–231, 233  
 Informed assessment of the risk, 534–536  
 Informed consent, 625, 894–896  
 Informed utilities, 535  
 Infrastructural high-tech projects, 827  
 Inherently safe design, 66  
 Inherent safety, 40–41, 43  
 Inherent utility, 585  
 Innovation, 1131, 1132  
 Innumeracy, 624, 629, 632–633, 637–650, 653, 654  
 Insensitive to scale, 687  
 Instinct for workmanship, 1142  
 Institutional design, 897, 901, 902, 904–905  
 Institutionalized individualism, 1022  
 Institutional risk managers, 249–251, 260, 261  
 Institutions, 882, 885, 893, 897, 901, 902, 904–905  
 Instrumental action, 1146  
 Instrumental rationality, 380, 1139, 1146  
 Instrumental reason, 1146–1148  
 Instrumental value, 326  
 Insula, 703  
 Integral emotions, 679  
 Integral feelings, 679, 682–683, 686–688  
 Integration, 1094, 1095, 1103, 1104, 1108, 1110–1112  
 Integrative approach, 186–187, 190–194, 198–200, 202, 203, 207  
 Intellectualization, 699, 700  
 Interaction, 762, 763, 765, 767–771, 780, 783, 784  
 Interactive-dependence, 77, 79  
 Interchange of equivalent parts, 528  
 Interdisciplinary, 1130, 1132–1134  
 Interdisciplinary research, 1094–1095  
 Interests, 180, 184, 185, 187, 189–196, 206, 208  
 Interfering/tampering with nature, 665, 666  
 Intergenerational equity, 297, 299–304, 307, 310, 311, 313–315  
 Intergenerational ethics, 934, 936, 937, 940, 943, 946, 957  
 Inter-generational impacts, 949  
 Intergenerational justice, 299, 301, 315, 933, 934, 941, 946, 954–956  
 Intergenerational risks, 932–958  
 Intergovernmental Panel on Climate Change (IPCC), 320–321, 323, 328, 330, 332, 615  
 Internal cardiac defibrillators (ICD), 493  
 Internal normativity, 71, 72  
 International Atomic and Energy Agency (IAEA), 297, 299, 301–304, 308–311, 315  
 International Risk Governance Council (IRGC), 1080, 1083  
 Interpersonal addition theorem, 434–437  
 Interpersonal relations, 706  
 Interpretation, 135–157  
 Interpretation of probabilities, 331  
 Intertemporal choice, 685  
 Interval scale, 413  
 Interval-valued functions, 550, 552  
 Intransitive preferences, 506  
 Intrinsic desire, 532–533  
 Intrinsic risk aversion, 526  
 Intrinsic utilities, 532–534, 556, 557, 560  
 Intrinsic-utility analysis, 532–534  
 Intrinsic value, 324  
 Intuition, 185–186, 198, 203, 204  
 Intuitive, 823  
 Investment game, 869, 870, 872  
 Investments  
   – return assessment, 527–528  
 In vitro, 495  
 In vivo, 495  
 Involvement, 701–703, 709–711  
   – level of, 709, 710  
 Iowa-gambling task, 824  
 IPCC. *See* Intergovernmental Panel on Climate Change  
 IPCC report, 323, 332  
 iPS-cells. *See* Induced pluripotent stem cells  
 Ipsilateral neocortex, 702  
 IRGC. *See* International Risk Governance Council  
 IRGC risk governance framework, 1081, 1084, 1085, 1087–1089  
 Irrational, 820, 825–828  
 Irrational choice, 483  
 Irresponsible conduct, 829  
 Irreversible environmental damage, 969

**J**

- Jackson case, 426, 427  
 Jasanoff, S., 1153, 1155  
 JCVI team, 492  
 Jeffrey, R.C., 381–382, 394, 398, 399  
 Jevons, S., 1144  
 Joint-stock company, 50  
 Jonas, H., 333, 965  
 Judicial documents, 966–967  
 Justice, 792, 797, 800–801, 821–822, 824, 827  
 Justification, 910, 917–918, 920, 922–929  
 Justified choice, 548

**K**

Kahneman, D., 1153, 1154  
 Kahn, H., 1143  
 Kaku, M., 1158  
 Kant, I., 793, 1144, 1145, 1147  
 KASAM, 299, 304–307, 311, 312  
 Kluver-Bucy syndrome, 700  
 Knightian uncertainty, 450, 605, 606  
 Knowledge, 916, 917, 922, 925  
   – limited, 514  
 Knowledge-ascription and practical stakes, 862

**L**

Land ethics, 326  
 Land use, 271–281, 283, 287, 290  
 Language, 695, 702–706  
   – community, 330  
 Laplace’s principle of insufficient reason, 330–331  
 Latent, 730, 731, 736, 738, 746  
 Lateralization of function, 702  
 Latour, B., 1021  
 Law, 1096, 1099–1100, 1112  
 Law of diminishing marginal utility, 378, 410–411  
 Law of large numbers, 64, 65  
 Law-social sciences, 1130, 1133–1134  
 Laws of probability, 479, 481  
 Lawyers, 1123, 1133  
 Layers of complexity, 763  
 Layers of values, 772  
 Laypeople, 665, 670, 820–822, 824, 826, 828, 830, 1007,  
   1009, 1017–1020  
 Layzer, D., 1158  
 Lead-time bias, 642–644  
 Learning, 584, 585, 594, 595  
 Left amygdale, 706  
 Left hemisphere (LH), 702–707  
 Left prefrontal cortex (Left PFC), 704, 706  
 Legal decision maker, 138, 144–146, 148, 149, 156, 157  
 Legal responsibility, 883–884  
 Leibniz, G., 1140, 1144  
 Leiss, W., 1148, 1149  
 Lenman, J., 911, 913, 917–921, 924  
 Leveling down objections, 911  
 Level of involvement (LI), 709, 710  
 Level of risks, 910–913, 920–923, 927–929  
 Levels of cognitive processing, 699–702  
 Levels of uncertainty, 87–109  
 Levitt, N., 1152, 1153  
 Lewontin, R., 1158  
 Lexicographically ordered set, 486, 488  
 Lexicographical ordering of options, 540  
 Lft PFC. *See* Left prefrontal cortex  
 LH. *See* Left hemisphere  
 LI. *See* Level of involvement  
 LI = (A + R)/2, 709  
 Liability, 882–884, 886  
 Liability-responsibility, 882  
 Liberal-conservative ideology, 741  
 Liberalism, 1059–1061  
 Liberal welfare approach, 893  
 Liberation of nature, 1146  
 Libertarian approach, 893, 894  
 Libertarian paternalism, 1148, 1154  
 Liberty, 193–195, 792–796  
 Life expectancy, 342  
 Lifestyle concerns, 1055–1056, 1058–1059  
 Lifestyle risks, 666, 670, 893  
 Lifeworld, 1148  
 Likelihood, 184, 195, 198, 201, 203, 205, 399  
 Likelihood ratio (LR), 138–148, 150–153  
 Limbic system, 704–705, 708–709  
 Limit, 181, 184, 186, 188, 192–195, 197–205, 207  
 Limited knowledge, 514  
 Limited liability, 50  
 Linear model, 164–165, 168–171, 174  
 Linguistic emotions, 704  
 1755 Lisbon earthquake, 351  
 Lithosphere, 322  
 Lobby, 827  
 Local communities, 766, 771, 773, 780, 782  
 Local knowledge, 1150, 1155  
 Locally global planning, 549  
 Locations of uncertainty, 92, 96  
 Logical positivists, 1141, 1146, 1147, 1150  
 Long-term risks, 932, 957  
 Loop of prerequisites, 781  
 Loss aversion, 583, 584, 586, 589  
 Lotteries, 384–387, 393, 397, 408, 413, 417, 418, 501, 502,  
   506, 511, 913–915, 922, 926, 927  
   – fair, 913, 926  
   – fairness, 914  
 Lotteries (horse), 482  
 Low-impact, high-frequency events, 226, 231  
 Low road, 701  
 LR. *See* Likelihood ratio  
 LR approach, 138–140, 144, 146, 148, 150, 156  
*Ludic Fallacy*, 445  
 Luhmann, N., 1004, 1010, 1014–1016

**M**

Mad cow, 1148  
 Magical thinking, 1150–1151  
 Major accident scenarios, 270, 275  
 Major actinides, 298, 309  
 Mammography screening, 625, 626, 629, 632–634, 640,  
   653, 655  
 Mannheim, K., 1142  
 Marcuse, H., 1146, 1148  
 Marginal utility, 115, 117, 119, 123, 125, 127  
 Marginal utility of money, 479, 480

- Marxism, 1142  
 Marxism–Leninism, 1156  
 Marx, K., 1142, 1146  
 Matching pennies, 592  
 Match of DNA, 141–144, 149–151, 155, 157  
 Mathematical expectation, 408, 427  
 MAX-ENT principle, 399  
 Maximal expected utility, 478  
 Maxi-max rule, 489, 551  
 Maximin rule, 379, 398, 971–972, 974  
 Maximization, 547, 548, 550, 567  
 Maximization for acts, 547  
 Maximize expected utility, 479, 486, 493  
 Maximize the minimal expected utility (MmEU), 487, 489  
 Maximizing objection, 546–550  
 Maximum expected monetary value-rule (MEMV),  
     378, 379  
 Maximum expected utility-rule (MEU), 379–382,  
     393, 400  
 Mayan civilization, 324  
 McDowell, J., 850  
 McIntyre, A., 835  
 Mean-preserving spread (MPS), 120, 121  
 Mean-risk formula, 527  
 Means-ends rationality, 380, 400  
 Measurement of utility, 530  
 Measures of risk, 526, 527  
 Mechanisms, 223–226, 229–231, 726, 727, 739–753,  
     755, 756  
 Media, 773  
 Media coverage, 747  
 Medical technology, 1053, 1055, 1058, 1061  
 Medical treatment, 796  
 The Medina Ortega Report, 1121  
 Member States, 1120, 1122, 1127, 1129, 1131  
 Memory, 681–683  
 MEMV. *See* Maximum expected monetary value-rule  
 Meno’s paradox, 99  
 Mental work, 700  
 Merton, R.K., 1151  
 Metacognition, 613  
 Methodological decisions, 889, 890  
 MEU. *See* Maximum expected utility-rule  
 Micro-economics, 500, 1144  
 Mill, J.S., 50, 792–793  
 Mindfulness, 697, 715, 717–718  
 Minimal risk, 186, 188, 192, 199–204, 207  
 Minimax, 551  
     – rule, 334  
     – strategy, 415  
 Minimized, 181, 185, 188, 189, 191, 192, 197  
 Minor actinides, 298, 308–311  
 Misattribution, 686–688  
 Mismatched framing, 627, 638  
 Misunderstandings, 762, 763, 773  
 Mixed groups, 564, 565, 571  
 Mixture space, 384  
 MmEU. *See* Maximize the minimal expected utility  
 Model building, 219–221, 232  
 Model design, 220, 229  
 Modeled encounters, 216–221, 230–233  
 Modeling, 213–235  
     – process, 219, 220, 231, 232  
     – techniques, 216, 217, 219–221, 232, 234  
 Model uncertainty, 98, 100, 102  
 Molecular manufacturing, 162, 175  
 Monderman, H., 262  
 Monetize, 797, 801  
 Money has diminishing marginal goodness, 409  
 Money has diminishing marginal value, 423  
 Money pump, 422, 503, 541–542  
 Monotonically increasing utility, 528  
 Monsanto, 1124–1126, 1128  
 Monty Hall problem, 508, 514  
 Moods, 678–680, 682, 686–688  
     – negative, 698  
     – positive, 698  
 Moral, 680, 687  
     – considerations, 825–827  
     – demands, 902  
     – expectations, 867, 868  
     – hazard, 452, 459  
     – responsibility, 164, 168, 170–175, 879, 883–884,  
       886–888, 897, 898, 902–903, 905  
     – rightness, 380  
     – traditions, 182  
     – values, 165, 167, 168, 171  
 Moralistic trust, 864  
 Morally salient aspects of risks, 820, 830  
 Morgenstern, O., 333, 382  
 Mortality rates, 623, 642, 644, 654, 655  
 Mosca, G., 1149  
 Motivation, 679–683, 687, 726, 740, 742, 743,  
     751–752, 828  
 Motivator, 680  
 Motives, 821–822  
 MPS. *See* Mean-preserving spread  
 Multi-attribute utility analysis, 519–522, 530, 531, 536, 537  
 Multi-attribute utility theory, 520, 522, 523, 537, 540  
 Multidimensional utility analysis, 522  
 Multiple utility scales, 540  
 Multi-step decision procedures, 540  
 Multivalued measures, 67  
 Mumford, L., 1151  
 Murphy, C., 983, 985–990, 992–995  
 Mutagenic, 810  
 Mutuality principle, 119  
 Mutual preferential independence, 531  
 Myopic loss aversion, 589  
 Myrdal, G., 1142

**N**

Nanodivides, 160, 161, 164, 175  
 Nanoelectronics, 163  
 Nanoparticles, 160–163, 172–173, 175, 490–492, 495, 496  
 Nano-safety, 490  
 Nanoscience, 493, 496  
 Nanotechnology, 34, 159–176, 478, 490, 492, 496, 736, 743–746, 825, 1072, 1086–1088  
   – risks, 743  
 Nanotubes, 490–492  
 Nanowires, 490–492  
 Narrative (anecdotal) evidence, 634  
 Narratives, 221, 233  
 Narrow bracketing, 586  
 Nash equilibrium, 538, 583, 584, 590, 592  
 National Academy of Sciences, 748, 749  
 National population, 780  
 Natural background, 33  
 Natural concept, 70–77, 82  
 Natural frequencies, 630, 639–641, 653–655  
 Natural gas, 774  
 Naturalists, 72–74  
 Naturalness, 33, 1053, 1056–1058, 1060, 1062, 1064  
 Natural rights, 923–924  
 Natural risks, 342, 878, 885, 886  
 Nature of risk and safety, 57, 58, 70  
 Nature of uncertainty, 92, 93  
 Nazi experiments, 182  
 N-body problem, 495  
 NEA. *See* Nuclear Energy Agency  
 Negative individualist, 709, 712  
 Negative moods, 698  
 Negative outcomes, 478, 490  
 Negative prosocial, 709, 712–714, 716  
 Negative prosocial emotions, 712, 713  
 Negative social value, 208  
 Negligence, 882, 885, 886  
 Negotiated exchanges, 871, 872  
 Neocortex, 700–702, 708  
 Net research risks, 196–197  
 Net risks test, 187, 189, 196–200  
 Neurobiological determinants of emotion, 696  
 Neurological research, 823  
 Neurotoxicants, 811, 812  
 Neutral evidence, 144  
 New age, 664  
 New Atlantis, 1140  
 Newcomb's problem, 508, 510, 511, 515  
 New Deal, 1142  
 New risks, 215, 235  
 NGOs, 343, 344, 348, 367  
 NIMBY (syndrome), 268, 269, 284–286, 289–291  
 Nobel Prize, 496  
 Nomothetic, 1147

**O**

Nonanthropocentric ethics, 957  
 Non-calculative conception of trust, 866, 874  
 Noncognitive decision, 546, 552  
 Non-event, 773, 781  
 Nonexpected utility theory, 116  
 Non-identity problem, 933, 937–944, 946  
 Nonlinear, 327–328  
 Nonlinear evolution equations, 321, 326  
 Non-point-source pollution, 794  
 Non-productive mischief, 795  
 Non-therapeutic procedures, 186–187, 191–192  
 Normalized frequencies, 640  
 Normative, 1096–1097, 1102  
 Normative concepts, 70, 71, 74–77, 82  
 Normative expectations, 867  
 Normative falsification, 504  
 Normative intuitions, 500, 503, 504, 508, 513  
 Normative judgment, 184, 185, 198  
 Normative principle of additivity, 531  
 Normative research, 500  
 Normative underpinnings, 973, 974  
 Normative validity, 500, 502–506, 508–510  
 Notion of risk, 913  
 Nozick, R., 44–46, 792–795, 912, 914, 924, 925, 927  
 Nuclear energy, 820, 827  
 Nuclear Energy Agency (NEA), 297, 299, 301–304, 308, 311, 313, 315  
 Nuclear fuel cycle, 297–299  
 Nuclear power, 663–667, 669, 736, 737, 742, 745, 747–749, 753, 754, 796, 797, 801, 802  
 Nuclear priesthood, 1151  
 Nuclear reactor, 827  
 Nuclear waste, 42, 662–664, 668  
 Nuclear waste management, 296–304, 308, 310, 314  
 Numerical probability estimate, 633, 635, 636  
 Nuremberg Code, 182, 202–203  
 Nussbaum, M., 844, 846, 847, 853, 981, 990

- Option's risk, 518, 526, 533  
 Orbitofrontal cortex (OFC), 701, 706  
 Ordering assumptions, 481  
 Ordering principles, 489  
 Ordinal scale, 412, 413, 427  
 Ordinal utility, 116, 384  
 Ordinal utility function, 411  
 Organizational responsibility, 883–884, 886  
 Organized irresponsibility, 897, 902, 1087–1089,  
   1120, 1121  
 Orgasm, 703, 704, 707  
 Origins, 1094–1109, 1112  
 Outcome, 376, 377, 379–391, 394–398, 406, 410–423,  
   425–433, 435–437  
   – risks, 490, 494  
   – uncertainty, 92, 101  
 Outcome, shared, 557–558, 565–566  
 Overdiagnosis, 641  
 Overdiagnosis bias, 642–644  
 Overlapping consensus, 903  
 Overtreatment, 626, 629, 641  
 Ownership, 50  
 Oxytocin, 872
- P**
- PACTS, 246  
 P-admissible, 486–487  
 Panic, 667  
 Paradigm, 1095–1097, 1103–1105, 1107, 1108, 1112  
 Paradigm shift, 815  
 Paradoxes, 248–260, 478–497  
   – of Ellsberg and Allais, 482–485  
 Parameters, 91, 92, 96–103, 106, 108–109  
 Parameter uncertainty, 321  
 Paretian preferences among options, 531  
 Pareto  
   – improvement, 411  
   – optimality, 49  
   – Principle, 431, 432  
 Pareto, W., 1149  
 Parkinson's disease, 491  
 Partial order, 552, 553, 555  
 Participation, 49, 51, 1073–1074, 1080, 1081, 1083  
 Participation process, 766–768, 779  
 Participatory, 1074, 1075, 1080, 1083–1085, 1087, 1088  
   – approaches, 1071, 1072, 1074, 1075, 1079,  
   1081–1083, 1085, 1086, 1089  
   – decision-making, 975  
   – TA, 1080, 1081, 1083, 1085  
 Partitioning and Transmutation (P&T), 296–297, 308–314  
 Past probabilities, 37  
 Paternalism, 206  
 Paternalistic, 195, 206  
 PCE. *See* Plausibility-comparable expectation  
 PE. *See* Plausibilistic expectation
- Peer effects, 584, 585  
 Peltzman, S., 250  
 Perceived control, 666  
 Perception, 1030–1046  
 Perception of technological risk, 874  
 Perfect moral storm, 937  
 Perfluorinated substances, 812  
 Permissible distributions, 486  
 Permissible utility functions, 486  
 Permissive decision principle, 541–542  
 Persistent substances, 811, 812  
 Personal interests, 190, 193  
 Personal involvement inventory, 709, 710  
 Personal probability, 479  
 Pesticides, 806–807, 811–814  
 PFC. *See* Prefrontal cortex  
 PFC-damaged patients, 704  
 Philosophical Conceptions of Trust and Risk, 859–868  
 Philosophy  
   – of economics, 48–50  
   – of law, 801, 802  
   – political, 29, 50–51  
   – of science, 37–40  
   – of technology, 40–43, 1051, 1061  
 Phobias, 825  
*Phronesis*, 847, 849, 854  
*Phronimos*, 837, 844  
 Physical probability, 526  
 Physicist, 1141, 1143, 1151, 1152, 1158  
 Phytoplankton, 322  
 Piecemeal social engineering, 1143  
 Pigou, A., 49  
 Placing trust, 864, 869, 872  
 Plato, 1139, 1140, 1144  
 Plausibilistic expectation (PE), 553–556, 559–561, 564,  
   569–571  
 Plausibility, 547, 552–555, 557, 559–561, 563–565, 569  
 Plausibility-comparable expectation (PCE), 556, 560,  
   561, 564, 565, 571  
 Plausibility-comparable groups, 564  
 Plausibility proof, 1123, 1124, 1126–1130  
 PMH. *See* Problem of Many Hands  
 Poisson distributions, 359  
 Poisson processes, 351, 359  
 Polarization, 739, 742–746, 752  
 Policy, 215, 217, 218, 226, 227, 229–233  
 Policy E, 919  
 Policy F, 919, 920  
 Policy makers, 820, 826, 827  
 Policy-making, 217–218, 231, 662, 670  
 Policy scientists, 1133  
 Political  
   – deficit, 1129–1130  
   – philosophy, 29, 50–51  
   – sciences, 1099, 1101

- Politicians, 662, 670–672  
Politicization of uncertainty, 1128–1129  
Pollutants, 806, 814, 815  
Polluter pays, 955  
Polluter Pays Principle (PPP), 900  
Pollution  
– non-point-source, 794  
Polywater, 34  
Popper, K., 1143, 1151  
Population growth, 342, 348  
Population management, 361  
Population-simulation model, 224, 225, 227  
Populist pitfall, 828, 829  
Portfolio theory, 50  
Positive individualist, 709, 714  
Positive individualistic emotions, 713, 716  
Positive linear transformation, 480  
Positive moods, 698  
Positive prosocial emotions, 709, 712–714, 716  
Positive risk, 453  
Positivism, 1141, 1146, 1147, 1151, 1152  
Positivistic risk approach, 1096–1097  
Positivistic risk paradigm, 1096, 1104–1105, 1112  
Possible worlds, 533, 538  
Posterior odds, 140, 142–144, 146, 148  
Posterior probability, 139–141, 143, 146, 399  
Post-industrial society, 1143, 1149  
Postmarket laws, 806–811, 813–816  
Postmarket risk assessments, 806, 809, 813–815  
Post-traumatic stress disorder (PTSD), 705  
Potency assessment, 808  
Potential harm, 184  
Potential social benefit, 181, 186, 189, 192, 193, 196, 198–200, 202–204  
Powers, 1002, 1006, 1010, 1012, 1014, 1015, 1018–1021, 1023  
Practical implications, 199–200  
Practical rationality, 820, 827  
Pragmatic arguments, 422  
Pragmatic justifications, 503  
Precaution, 911, 915, 917–929, 1099–1100, 1108, 1111  
Precautionary approach, 321, 333–337, 813, 966–968, 975  
Precautionary measures, 911, 917–921, 928, 929  
Precautionary principle, 31, 39, 496, 669, 802, 803, 933, 946, 953, 962–975, 1122–1124, 1132, 1133  
– strong formulation, 334  
– weak formulation, 334  
Precautionary response, 969, 970  
Precautionary saving, 122, 123  
Precautionary thesis, 918–921  
Precaution-based regulation, 1123  
Precision objection, 546–547, 550–552  
Predictions, 216–234  
Predictive expectations, 867  
Preface paradox, 381  
Preference consistency, 501, 503  
Preferences, 383–387, 389–397, 399–401, 519, 520, 522–523, 526, 529–532, 539, 797–800, 802  
Preferences-as-memory framework, 682  
Preference satisfaction theory, 412, 435  
Preferential independence, 531  
Prefrontal cortex (PFC), 701, 704–707  
Prejudices, 825  
“Pre-live” emotional consequences, 716  
Premarket risk assessment, 814–816  
Premarket testing, 807, 814–816  
Premarket toxicity testing, 806, 811–816  
Preparation, 770, 782  
Preparedness, 344, 354, 357, 367  
Prevention principle, 969  
Prevention probabilistic approach, 276, 278  
Preventive medicine, 219  
Primary, 1013  
Primary and reflexive scientification, 1009, 1013  
Primary appraisal, 700  
Primary prevention, 40  
Primitive, 1150, 1155  
Primitive evaluative perception, 700  
*Primum non nocere*, 965  
Principles  
– of compensation for risk, 795  
– of comprehensive rationality, 548  
– of diminishing responsibility, 297, 304–308, 313  
– of expected utility for decisions, 548  
– of expected value, 408, 409  
– of independence of irrelevant alternatives, 489  
– of indifference, 399  
– of insufficient reason, 399  
– of justice, 305  
– of mathematical expectation, 408  
– of non-exploitation, 196  
– of personal good, 431–437  
– of ratification, 538  
– of utility, 547, 548, 550  
– of utility maximization for decisions, 547, 548, 550  
– of utility satisfaction for decisions, 550  
Prior explicit consent, 794  
Priority view, 437  
Prior odds, 140, 142–144, 146–148, 150  
Prior probability, 139, 143, 146, 399  
Prisoners’ dilemma, 511–513, 866, 869  
Prisoner’s dilemma, 869  
Privacy, 160–163, 174, 175  
Private, 1052, 1053, 1056–1059, 1062–1064  
Prizes, 384–387, 392, 393, 397  
Probabilistic (approach to risk prevention), 276, 278  
Probabilistic beliefs, 502

- Probabilistic mixtures, 44, 45, 47  
 Probabilistic safety analysis (PSA) results, 772  
 Probability, 30, 31, 35–37, 41–43, 45, 46, 51, 56, 58, 59, 61, 63–73, 76, 81, 82, 136–150, 152, 153, 156, 157, 376–382, 384–391, 393–396, 398–401  
   – agreement theorem, 530  
   – of an unwanted effect, 821  
   – density, 147, 152  
   – distributions, 359  
   – laws of, 479, 481  
   – limit, 46  
   – measures, 417, 420, 421  
   – neglect, 685, 822  
   – numerical estimate, 633, 635, 636  
   – past, 37  
   – prior, 139, 143, 146, 399  
   – theory, 35  
 Problematic communication aspects, 763, 773–774  
 Problem-focused, 700  
 Problem framing, 1071, 1075, 1085, 1089  
 Problem of Many Hands (PMH), 878, 879, 892, 897–905  
 Problem of paralysis, 923–924  
 Procedural requirement, 971, 972  
 Procedure of responsibility distribution, 897, 901–904  
 Process fairness, 582–583  
 Production of risks, 1012, 1022, 1024  
 Productive activity, 795, 796  
 Product transitivity, 564  
 Profession, 887, 888, 903  
 Professional integrity, 196, 206  
 Proportionality of individual risk and potential social benefit, 205  
 Propositional attitudes, 394  
 Propositions, 552, 553, 557, 558, 568  
 Proposition's intrinsic utility, 532–533  
 Prosecutor's fallacy, 143  
 Prosocial behavior, 871, 873  
 Prosocial emotions, 711  
 Prospect theory, 116, 485, 488, 507–508, 580, 583  
 Prostate-specific antigen (PSA)  
   – screening, 624, 625, 627, 629  
   – tests, 624, 625, 627, 628  
 Protection, 912, 922, 924, 925, 928  
 Protest group, 861–863  
 Prudence, 115, 120–127, 129, 130  
 PSA. *See* Probabilistic safety analysis results  
 Psychological approach, 56, 57, 70, 77  
 Psychological mechanisms, 726, 727, 872  
 Psychometric, 730, 731, 733, 737, 739  
 Psychometric model, 662, 664–667, 670, 672  
 Psychometric school of risk analysis, 1006  
 Psychometric studies, 821  
 Psychometric theory of risk, 739  
 Psychophysical numbing, 684–685  
 P&T. *See* Partitioning and Transmutation  
 PTSD. *See* Post-traumatic stress disorder  
 Public, 820, 821, 825–30, 989–991, 993–995  
 Public concerns, 1070–1072, 1075, 1076, 1082, 1083  
 Public confidence, 196, 206  
 Public controversies, 1070, 1074, 1082, 1083, 1085  
 Public debates, 820, 829, 830  
 Public dialogue, 1052, 1053, 1064, 1065  
 Public discontent, 1052  
 Public engagement, 1083  
 Public good games, 596  
 Public health, 750–752, 962, 965  
 Public health and safety appeals, 695, 711–712  
 Public health risk, 213–235  
 Public risk perceptions, 821  
 Pure intergenerational problem, 933, 935–937  
 Pythagoreans, 1139
- Q**
- Qualitative differences in the source of risk, 797  
 Qualitative dimensions of risk, 797  
 Qualitative probability, 419  
 Quantifiability, 1059–1063  
 Quantifiable objects, 218  
 Quantitative, empirical information, 827  
 Quantitative methods, 822, 824  
 Quants, 1151–1152  
 Quantum field theory, 495
- R**
- Race, 741, 742, 745  
 Radiotoxicity, 296, 298, 308, 312  
 RAG, 446, 467–472, 474  
 RAM. *See* Reason-Affect Map  
 Ramsey, F.P., 387–391, 394, 398  
 Random match probability, 141–143, 149, 157  
 Rank-dependent expected utility, 507–508  
 Rankings of options, 524  
 RAPs. *See* Reason-affect profiles  
 Rational  
   – agent, 376, 380, 383–390, 393, 395, 396, 398, 399  
   – choice, 489, 493  
   – cognitive reactions, 698  
   – decision-making, 478–483, 490, 494, 496  
   – decision theory, 821  
   – individual, 407, 410–412  
   – involvement, 709–711  
 Rationalists, 1140, 1144, 1145  
 Rationality, 481, 482, 485, 494, 501, 503, 506, 508, 512–514, 519, 520, 526, 538, 541–542, 694–699, 702, 707, 709–711, 714, 715, 717, 1139, 1140, 1143–1150, 1154, 1155, 1158  
   – ontological and structural axioms, 480, 482  
   – requirements, 424  
   – strategic, 874  
   – theories, 494

- Rationalization, 1146  
Rational Man, 481, 482, 490  
Ratio scale, 413  
Rawls, J., 46, 910  
Realism, 1004, 1016, 1020–1022  
Realist, 1100  
Realistic decision principles, 537  
Realistic decision problems, 537  
Realistic disaster scenarios, 452, 461  
Reality of risk, 79, 80  
Real-life decision, 545–571  
Real rationality, 549  
Real-world communications environment, 717  
Reason-affect map (RAM), 710  
Reason-affect profiles (RAPs), 711  
Reciprocal exchanges, 47, 871, 872  
Reciprocity, 869, 870  
Recklessness, 882, 885, 886  
RECs. *See* Research ethics committees  
Red Flag Act, 241  
Reducing risks, 884, 888  
Reductionism, 70, 74–76  
Reduction of complexity, 865  
Reduction of risk, 911, 913, 919–921  
Reference class, 915, 916  
Reference class problem, 150, 154  
Reference-dependence, 77, 78, 80, 580, 583  
Reference group, 765  
Reference point, 577, 580–584  
Referendum, 771, 772, 781  
Reflection, 824–826, 828, 1100–1101, 1108  
Reflection principle, 1095, 1111–1112  
Reflective emotional capacities, 824  
Reflective equilibrium, 503, 504  
Reflexive, 1010  
Reflexive modernization, 1004, 1012–1014  
Reflexive scientification, 1009, 1013  
Regret, 585, 678, 679  
Regulation, 679, 680, 688  
Regulation of inland waters, 322  
Regulators, 1120, 1121, 1130, 1133  
Regulatory decisions, 519, 536  
Reification, 445–447  
Relation
  - between feelings-based and cognitive processes, 681
  - between individual risk and potential social benefit, 193, 199, 204
  - between risk and responsibility, 878–879, 885–886, 905
  - between trust and risk, 859–862  
Relative disutility, 552, 556–558, 567  
Relative risk, 627, 632, 636–639, 646, 654, 655  
Relative risk aversion, 115, 118, 122, 125  
Relative safety concept, 61  
Relative utility, 556, 557  
  
Reliability, 222, 230–232  
Relief, 678  
Remote possibilities, 34, 51  
Repository for spent nuclear fuel, 771  
Representation and uniqueness theorems, 420–422  
Representation theorems, 382–387, 390, 394–395, 412, 480, 501, 520, 530–532  
Representative participation, 779  
Reprocessing, 296, 298, 303, 308, 310–312, 314, 316  
Reproductive rate, 227–229  
Reptilian, 708–716
  - emotions, 708, 714, 716
  - erotic emotions, 713
  - power emotions, 711, 713, 714  
Reptilian power, 711  
“Reptilian rewards” hypothesis, 712, 714  
Reptilian sex, 711  
Reputation, 595  
Requirement of rationality, 422  
Research ethics committees (RECs), 185, 186, 188, 194–196, 202, 203, 206, 207  
Research interventions, 182–189, 191–193, 197, 199, 203  
Research priority-setting, 205  
Research risks, 180, 181, 192, 193, 196–197, 199–205, 207  
Research risks “excessive”, 198  
Resolute choice, 582  
Resources, 981, 982, 984, 986, 990–993, 996  
Respect individuals, 917, 920  
Responsibility, 580, 581, 588–590, 595, 596, 661–673, 878–905, 1050, 1053, 1061, 1063–1065  
Responsibility as a virtue, 883, 886, 897, 901–902, 904, 905  
Restriction of individual rights and liberties, 795  
Retrievability, 297, 304, 310, 312  
Return assessment of investments, 527–528  
Revealed preference method, 798–799  
Reversal of the burden of proof, 968  
Reversibility, 1110  
Reviews, 762, 766, 782  
RH. *See* Right hemisphere  
RH hypothesis, 702, 703  
Rickert, H., 1147  
Right amygdale, 706  
Right hemisphere (RH), 702–705  
Right PFC, 704, 706  
Rights, 792–795, 911, 917, 923–924, 926  
Rights-based ethical theories, 795, 796  
Rights-based ethics, 923–924  
Rights-based moral theory, 45–47  
Rio Declaration of 1992, 802  
Rio formulation, 962, 970, 972  
Risk, 55–82, 179–208, 376–401, 518–542, 694–698, 708, 709, 712, 715, 806–816
  - absolute, 627, 628, 637–639, 652, 654, 655
  - acceptability of, 881, 882, 886, 893, 896, 899, 905
  - acceptable, 56, 57, 60, 61, 67, 71, 72, 74

- of accidental handgun shooting, 747
- analysis, 28–30, 37, 42, 48, 49, 51, 478, 494, 496, 497, 822, 1094, 1096, 1100–1101
  - Socratic approach, 494, 496
- as an attribute, 518, 523–527, 534, 536
- anonymous, 922
- appetite, 453, 455, 462, 467, 471, 474
- assessments, 214, 216, 217, 223, 230–233, 806–816, 880, 883, 886–890, 895, 897, 965, 968, 970–974, 1121, 1125–1128, 1130, 1131
- aversion, 114, 115, 117–118, 120–123, 125, 127–129, 261, 263, 480, 509, 708, 824
  - about goodness, 428, 429
  - about money, 408, 423, 428
- aversity, 396, 397
- awareness, 784
- catastrophic *vs.* chronic, 821
- characterization, 808
- classical interpretation, 1074–1079
- communication, 496, 621–655, 693–718, 762–785, 883, 887, 894–897
- comparisons, 192, 201, 202, 207
- compensation, 250, 251, 255, 795, 927
- consent, 920, 923, 927
- controllers, 494
  - of daily life, 199, 201, 202
  - of death, 794, 796, 798, 800, 801
  - definition, 913
  - denial, 669
  - determination, 980, 984–989, 991, 992
  - distribution, knowledge, 922
  - distribution, 915, 916
  - emotions, 820, 822, 823, 825, 826, 828–830
  - evaluation, 860–862, 865, 868, 980, 984, 989–991, 994–996
  - exchange, 924
  - as feeling, 822
  - to future generations, 802
  - *vs.* goods, 912
  - governance, 219, 223, 235, 1003, 1004, 1016–1017
  - homeostatis, 452
  - imposition, indiscriminate, 915–917, 922
  - impositions of, 922
  - indiscriminate, 915–917, 922
    - individual, 193, 199, 204, 205
    - intelligence, 603–619
    - level of, 910–913, 920–923, 927–929
    - of lifestyle, 666, 670, 893
    - limits, 186, 192, 193, 199–203, 207
    - longterm, 932, 957
    - management, 494, 496, 881–883, 887–891, 895–897, 901, 963, 968, 972, 974, 980, 984, 989, 991–993, 995, 996, 1002, 1003, 1006, 1009, 1011, 1017, 1019, 1020, 1022–1024, 1120, 1121, 1125, 1127–1128, 1131
    - management strategy, 776
    - manager, 1122, 1129–1130
    - meaning, 913
    - measure, 3
    - messages, 695, 696, 698, 699, 712, 715–717
    - minimal, 186, 188, 192, 199–204, 207
    - mitigation, 667, 669
    - nature of, 57, 58, 70
    - neutrality, 429
      - neutrality about good, 428, 429, 433
      - neutrality about money, 408, 414
    - notion, 913
    - option, 518, 526, 533
    - outcome, 490, 494
    - perception, 661–673, 726–729, 734–744, 746, 750, 753, 755, 880, 881, 885, 1030, 1032–1046, 1097, 1098, 1101, 1102
    - perception studies, 58, 67
    - perspectives, 1101, 1109, 1110
    - philosophical conceptions, 859–868
    - policy makers, 827
    - politics, 218–219, 826–829
    - positive, 453
    - positivistic, 1096, 1104–1105, 1112
    - *vs.* precautionary measures, 920
    - prevention, 276, 278
    - producer, 1124, 1128–1130
    - proneness, 480
    - protestors, 1128–1130
    - psychometric, 739
    - public, 821
    - rationality, 215, 231, 234, 235
    - reduction, 883, 886–888, 893, 895, 897, 911–913, 917–921, 926, 928, 929
    - reduction, balancing of, 911
    - reduction *vs.* fairness, 926
    - relative, 627, 632, 636–639, 646, 654, 655
    - research, 1122–1134
    - and responsibility, 878–879, 885–886, 905
    - scholars, 1120, 1133
    - seekingness, 824
    - sensitivity, 665, 671–673
      - of serious/lasting harm, 184, 201, 203
    - as side-effect, 925
    - society, 88, 93, 95, 96, 100, 108, 342, 347, 363, 881–882, 885, 886, 1003, 1004, 1012–1014, 1022, 1023, 1149, 1157
    - society acceptability of risks, 881–882, 885, 886
    - source of, 797
    - spreading, 50
    - taking, 29, 43, 44, 48–50, 52, 497, 872
    - technical conception, 880–882, 885, 889
    - technological, 342, 728, 740–741, 743, 802, 874, 879, 884, 886, 887, 894, 904–905
    - from terrorism, 829

- theory of, 725–757
  - thermostats, 250, 257–263
  - tolerance, 120, 129
  - as-value, 681
  - voluntary, 928
  - voluntary risk taking, 928
  - weighing, 49
  - “Risk-based” system of research oversight, 207
  - Risk-benefit
    - analysis, 48–49, 821, 822
    - calculation, 678
    - profile, 179–208
  - Risk communication model of transparency (RISCOM), 762, 764
  - Risk-cost-benefit analysis preferences, 797
  - Risks Outside the Box (ROB), 451, 454, 458, 461, 467, 468, 470, 473
  - Risky shift, 580
  - Risky technologies, 820, 826–830
  - RMS, 452, 457, 458
  - Road Safety Act of 1967, 243
  - Road to failure, 766
  - ROB. *See* Risks Outside the Box
  - Robbins, L., 49
  - Role responsibility, 882, 905
  - Romantic desire, 708
  - Romantic movement, 1144
  - RoSPA, 246
  - Roulette lotteries, 387
  - Routes of exposure, 811
  - Routine examinations, 201
  - Row disutility, 567
  - Rule of choice, 971–972
  - Russian roulettes, 794–796, 912, 914, 915, 922
- S**
- Safer sexual behavior, 712–715
  - Safe sex communication scale (SAFECOMM), 712, 714
  - Safety, 30, 31, 40–43, 45, 51, 52, 879, 880, 883, 887, 888, 891, 893, 896, 898, 899, 905
    - barriers, 41–43
    - belts, 45
    - data sheets, 776–777, 783
    - engineering, 40, 43, 51
    - factors, 40, 41, 43
    - nature of, 57, 58, 70
    - obsessed, 783, 784
    - as spatial value, 286–290
  - Salmon, W., 1158
  - Salomon’s House, 1140
  - 1906 San Francisco earthquake, 351
  - Satisficing, 546, 549, 550, 553, 554
  - Savage, L.J., 391
  - Scandinavian Myth, 244–245, 261
  - Scanlon, T.M., 911, 917, 918, 923–925
  - Scarcity, 708
  - Scenario building, 216, 217, 221–223, 226, 227, 229–231, 233
  - Scenarios, 148, 149, 220–222, 226, 228–231, 234
  - Schelling, T., 798, 800
  - Schllick, M., 1150
  - Science, 160, 164–172, 176, 1120–1121, 1123, 1124, 1126–1128, 1130–1134
    - communication, 752
    - in Society, 784
    - and technology studies (STS), 1094–1096, 1098–1099, 1112, 1133
    - of trust as a moral attitude, 870, 872–873
    - of trust as non-calculative, 871–872
    - wars, 1152–1153
  - Scientific
    - approaches to trust and risk, 868–873
    - consensus, 747–749
    - corpus, 37, 38
    - education, 1130
    - value, 205
  - Scientification of politics, 1130
  - Scientisation, 1003, 1006, 1007, 1009, 1010, 1013, 1017–1021, 1024
  - Scientist approach, 56, 57, 70, 71, 74
  - Scientization
    - reflexive, 1009, 1013
  - Scope insensitivity, 684–685
  - SEA. *See* Strategic environmental assessment procedure
  - Seatbelt, 241, 244–249, 251–252, 255, 260
  - Secondary appraisal, 700
  - Secondary prevention, 40
  - Secondary uncertainty, 457, 458, 463
  - Second-order probabilities, 37, 487
  - Security, 40, 50
  - Security-optimal, 486–489
  - Seismic hazard map, 346
  - Selection of attributes, 523–524
  - Self-affirmation, 752–753
  - Self-interest, 826, 827, 829
  - Selfish emotions, 703–705, 709, 713
  - Selfishness, 512
  - Self-preservation, 705, 709
  - Self-serving interpretations, 590
  - Self-torturer, 506, 515
  - Sen, A., 981, 982, 988, 990, 994
  - Sense-dependence, 78, 79
  - Sense of urgency, 827–828
  - Sensitive individuals, 33
  - Sensitivity analysis, 328, 538
  - Separability, 536, 537
    - of attributes’ utilities, 537
    - of moral values, 537
  - Sequential-analytic, 699, 702

- Sequential prisoner's dilemma, 869  
 SERR method. *See* Systematic evaluations of research risks method  
 Set of attributes, 519–524, 528, 531, 532, 537  
 Set of principles, 1095, 1108, 1112  
 Seveso Directives, 273, 275, 278  
 SEVESO II Directive, 767, 774  
 Sex differences, 706–707, 712  
 Sexual norms, 752  
 Shared decision making, 625, 629, 650, 654  
 Shared outcome, 557–558, 565–566  
 Sharpe ratio, 528  
 Shockley, W., 1157  
 Short-circuited, 699–700  
 Sidgwick, H., 793  
 Silent majority, 670  
 Similarities, 771, 777, 778  
 Simple risk problem, 1071, 1080, 1081  
 Simulation models, 217, 222–225, 227–230, 232  
 Simulation program, 220, 221  
 Single-attribute evaluation, 522  
 Single-event probability, 640, 641, 655  
 SINTEF/NTNU in Trondheim, 774  
 Site-specific risk, 267–270, 284, 287, 289  
 Slash and burn, 322  
 Slovic, P., 821, 822, 1153  
 Small world, 416  
 Smeed law, 260  
 Smith, A., 49–50  
 Snowballing-uncertainty effect, 350  
 Social
  - amplification of risk, 1006–1009
  - Amplification of Risk Framework (SARF), 1098, 1100
  - benefit, 193, 199, 204, 205
  - brain, 705
  - capital, 783
  - construction of risk, 1006, 1008, 1009
  - constructions, 57, 70, 77–80
  - constructivism, 1152
  - contract, 1156
  - emotions, 704, 706
  - engineering, 1139, 1142, 1143
  - facilitation, 580
  - functions, 230
  - interaction, 706
  - loss aversion, 583
  - loss frame, 583, 584
  - marketing, 624
  - meaning overdetermination, 755, 756
  - models, 164–168, 170, 171, 173
  - perspective, 800, 801
  - preferences, 529, 530
  - production of risk, 1012, 1022
  - proof, 708
- psychology, 580, 586, 595, 597  
 – reference points, 577, 580–584  
 – regret, 577, 585  
 – risk-taking, 870  
 – roles, 1032, 1038, 1040, 1041  
 – system, 1148  
 – trust, 668  
 – value, 181, 184, 185, 189, 191–195, 197, 198, 203–206, 208  
 Sociology, 1097, 1098  
 Socrates, 28, 1139  
 Socratic approach, 495, 496
  - to decision-making, 478
  - to risk analysis, 494, 496
 Soft impacts, 1049–1065  
 Solidaristic, 727, 728  
 Somatic marker hypothesis, 701, 707  
 Somatic markers, 683  
 Sorites paradox, 505, 506  
 Sources of uncertainty, 93, 109  
 Soviet of engineers, 1142  
 Spatial dimension (of technological risks), 285  
 Species preservation, 705, 709  
 Spent fuel, 298, 299, 304, 307–310, 313, 314  
 Spinoza, B., 1140, 1144  
 Spontaneous emotional communication, 703  
 Spotlight, 680  
 Stable basic goals, 538  
 Stag hunt game, 590, 591  
 Stakeholder participation, 1120, 1132, 1133  
 Stakeholders, 670  
 Stalemates, 829  
 Standard deviation, 527, 528  
 State, 407, 408, 411, 412, 414–423, 425–427, 430–437  
 State-dependent utilities, 482, 488  
 State of nature, 115, 116, 118, 119, 123, 125, 127, 129, 415, 482, 486, 487  
 Statistical evidence, 633–634, 637  
 Statistical expectation value, 879, 880  
 Statistical illiteracy, 624, 637–650  
 Statistical literacy, 627, 630–632, 637, 650, 652–655  
 Statistical-probabilistic, 214, 215  
 STDs, 750  
 Stem cell research, 478, 485, 490, 492  
 Steps in a decision process, 784  
 Stereotypes, 825  
 Stigmatization, 496  
 Stigmatized, 494, 496  
 Stochastic dominance, 121  
 Stochastic reference points, 584  
 Stochastic uncertainty, 272  
 St. Petersburg paradox, 378, 379, 396, 409  
 Strategic environmental assessment (SEA)
  - procedure, 768
 Strategic rationality, 874

- Strategic trust, 864  
Strategic uncertainty, 590–592, 594  
Strategies for risk communication, 778  
Stretch, 765  
Structural axioms, 480, 482  
STS, 1094–1096, 1098–1099, 1112, 1133  
Subcortical estimative system, 699  
Subjective probabilities, 415, 423, 425, 427, 431, 522–523  
Subjectivity, 63–66, 68, 70, 72, 77, 79–81  
Subpopulations, 33  
Sunk costs, 542  
Sunstein, C.R., 845, 846, 1148, 1153, 1156, 1159  
Sure-thing, 483, 488  
Sure thing axiom, 392  
Sure-thing principle, 417–419, 422, 423, 430, 431, 435, 437, 481–484  
Surface storage, 299, 313, 314  
Surveillance, 868  
Survival rates, 637, 641–644, 654, 655  
Swimming pool drowning, 747  
Sympathy, 824, 825  
Syncretic cognition, 702, 705, 709  
Synthetic biology, 478, 492  
System 1, 823, 826, 828, 1154  
System 2, 823, 828, 1154  
Systematic evaluations of research risks (SERR) method, 202  
Systematic thinking, 698
- T**
- Take responsibility, 901, 902  
Taleb, N.N., 606  
Tang Dynasty in China, 324  
Target system, 321, 327–329, 331  
Technical conception of risk, 880–882, 885, 889  
Technical rationality, 215, 233–234  
Technical risk analysis, 1003, 1005–1009, 1017, 1018, 1020, 1024  
Technocracy, 1137–1160  
Technocracy, Inc., 1142  
Technocratic, 821, 826, 828, 829, 1124, 1129–1130  
Technocratic pitfall, 828, 829  
Technocratic rationality, 1144, 1158  
Technological development, 342  
Technological risks, 342, 728, 740–741, 743, 802, 874, 879, 884, 886, 887, 894, 904–905  
Technologies of governance, 221  
Technology, 160, 162–168, 170, 172, 174–176  
– assessment (TA), 1067–1090  
– based laws, 807, 809  
– based statute, 808  
Technostructure, 1143, 1159  
Teleological ethics, 424  
Teller, E., 1143, 1158  
Temperance, 115, 120–127, 130  
Temporal lobe, 704–705  
Terminological issues, 967  
Terrorism, 668, 671, 672  
Terrorism insurance, 458, 460  
Terrorists, 35, 40, 253, 254  
Testosterone, 872  
Texas Sharp Shooter fallacy, 155  
Text messaging, 240  
Thalidomide, 812, 813  
*The Black Swan*, 445, 461, 606  
Theory of mind (ToM), 705–706  
*The wisdom of crowds*, 670  
Thick concepts, 70, 74–77  
Thick normative concepts, 74, 77  
Third parties, 208  
Thoreau, H.D., 326  
Thought experimentation, 504  
Threat, 962, 964, 966, 969, 970, 973  
Three claims, 764  
Thyroid disruptors, 811  
Tiger attacking, 926  
Titanic, 42  
Titanic effect, 256–257  
ToM. *See* Theory of mind  
Tool box, 780  
Total ignorance, 92, 96  
Total outcome, 557, 558  
Toxicants, 806–808, 811–816  
Toxicity data, 810  
Toxicity of products, 806  
Toxic responses, 811  
Toxic Substances Control Act, 807  
Trade, 1122, 1131–1133  
Trade-offs between goals, 518  
1962 Traffic Act, 242, 244  
Traffic safety policy, 891  
Tragedy of the commons, 596, 898, 899, 904, 905  
Transitive, 411, 412, 417, 480, 481, 502  
Transitivity, 411, 412, 422, 481, 552, 562–565, 571  
Transparency, 542  
Transparency forum, 765  
Transparent evaluation of options, 518  
Triad of compensation, consent and level of risk, 927  
Trust, 51, 95–96, 98, 101, 102, 108, 490, 494, 496, 668, 673, 763, 764, 777–780, 782–784, 1003, 1004, 1011, 1016–1020, 1132, 1133  
– beyond rational calculation, 864–866  
– game, 592–594  
– honor game, 869  
– as a moral attitude, 860, 865–868, 870, 872–873  
– orientations, 869  
– philosophical conceptions, 859–868  
– placing, 872  
– and risk, 858–874  
– as a special instance of rational risk-taking, 863  
– strategic, 864

- Trustee decisions, 534–536  
 Trustworthiness, 860, 862, 863, 871  
 Tuskegee syphilis trial, 183  
 Tversky, A., 1153, 1154  
 Two cultures, 29  
 Two-envelope paradox, 511, 515  
 Two-step decision procedure, 541  
 Type II errors, 810  
 Types
  - of decision, 492, 493
  - of potential benefits, 206
  - of risk, 196, 206
 Typology of risk, 1095, 1104–1108
- U**
- Ultimatum game, 592  
 Uncanny, 825  
 Uncertain risks, 1120–1124, 1126, 1128, 1130, 1132, 1133  
 Uncertainty, 31, 41–44, 46, 47, 49, 50, 56, 59, 63, 65–69, 72, 73, 76, 80–82, 215, 217, 218, 221, 227–230, 406–408, 413, 415, 416, 418, 419, 422–429, 431–434, 436, 437, 623, 625, 629–633, 635, 637, 645, 647–650, 654, 764–766, 775–777, 782, 880, 888, 890, 894, 896, 897, 900, 964, 966, 967, 969, 972–974, 1071, 1075–1077, 1079–1081, 1084, 1085, 1087–1089, 1094–1096, 1099–1100, 1103–1112
  - about the model, 98
  - about the outcome, 97
  - about the parameters, 97
  - certain, 1107
  - conceptualization, 321
  - information, 1123–1126, 1129
  - intolerance, 1124–1126, 1128–1130, 1133
  - Knightian, 450, 605, 606
  - level of, 87–109
  - locations of, 92, 96
  - long-term, 300, 306–308, 310, 314
  - nature of, 92, 93
  - outcome, 92, 101
  - paradox, 1121–1124, 1128–1131, 1133
  - parameters, 101, 321
  - with parameters, 108–109
  - politicization, 1128–1129
  - sources of, 93, 109
  - strategic, 590–592, 594
  - surrounding choosing the model, 103
  - tolerant experts, 1130
  - training, 1130
 Unconditional utility function, 480  
 Understanding, 820, 824–826, 828, 829, 1141, 1145, 1147, 1150, 1153, 1160  
 Undue burdens, 297, 301, 313, 314  
 Unequal distributions, 913, 921–928
  - fairness, 922–927
  - justification of, 922–927
 Unethical actions, 829  
 UNFCCC. *See* United Nations Framework Convention on Climate Change  
 Unintended collective consequences, 904  
 Unintended communication, 765  
 Uniqueness theorem, 412, 420–422  
 Unique outcome, 557, 558, 565–566  
 Unique probability measure, 485–486  
 United Nations, 983  
 United Nations Framework Convention on Climate Change (UNFCCC), 326, 333–334  
 Unknown effects, 34  
 Unknown unknowns, 99, 102, 106, 349, 357, 358, 445, 446, 461–462, 467  
 Unobtrusive challenge, 763  
 Unreasonable demands for scientific studies, 810  
 Unscientific, 963, 973  
 Upper limits, 181, 184, 188, 192, 193, 197–200, 202–205, 207  
 Upper risk limits, 192, 193, 199, 202, 203, 207  
 Upstream, 1083  
 Uranium (U), 296–298, 300–304, 308–313, 315  
 U.S. Environmental Protection Agency, 807, 808  
 Utilitarian ethics, 1139, 1144  
 Utilitarianism, 44, 45, 48, 49, 52, 380, 381, 437, 522, 536, 537, 821–824  
 Utility, 376, 379–391, 393–397, 399, 401, 519–542, 982, 985–986, 991, 992
  - analysis, 519–522, 530, 534, 536, 537
  - cardinal, 116
  - collective, 536, 537
  - comparable expectation (UCE), 556, 560, 564, 565, 571
  - comparable groups, 564
  - of equality, 537
  - expected, 36
  - function, 478, 480–482, 484–487, 501, 502
  - maximization, 547, 548, 550
  - monotonically increasing, 528
  - multidimensional, 522
  - ordinal, 116, 384, 411
  - principles of, 547, 548, 550
  - relative, 556, 557
  - representation, 529, 532
  - satisfaction, 550
  - scale, 520, 530, 540
  - theory
    - expected, 579, 580
    - multi-attribute, 520, 522, 523, 537, 540
 Utils, 1144
- V**
- Vaccination, 750, 756
  - programs, 820
 Vaccines, 222–227, 230, 232, 235

- Vagueness, 972  
 Valence approaches, 707  
 Valence hypothesis, 702–703, 707  
 Value, 680, 682–685, 687, 688, 799, 801, 802, 822, 825, 1070, 1071, 1074, 1075, 1077–1081
  - differences, 480, 482
  - free, 890, 1147, 1157
  - free risk assessment, 29
  - holism, 537
  - of human life, 798, 799, 802
  - judgments, 169, 411, 889, 890
  - ladenness, 170
  - neglect, 686
  - neutral, 168, 169
  - at Risk (VaR), 446, 452, 464, 1152
  - scales, 664
  - uncertainty, 490, 494
 Valuing life, 433–434, 436  
 Variability, 635, 647–648  
 Variance, 321, 526, 527  
 Variation in risk judgments, 186  
 Veblen, T., 1142, 1143  
 Vegetation period, 324  
 Ventromedial prefrontal cortex (VMPFC), 701, 706, 707  
 Verbal probability, 633, 635, 636  
 Veridical paradox, 504, 508, 509, 511, 513, 514  
*Verstehen*, 1147  
 Veto, 46  
 Victims, 826, 829  
 Videos, 694, 716–717  
 Vienna circle, 1141, 1150  
 View of trust, 863  
 Virtual risk, 248, 249, 251, 256–257, 260  
 Virtue, 883, 886, 888, 897, 901–902, 904, 905
  - epistemology, 337
  - ethics, 822, 826, 888, 897, 901–902
  - responsibilism, 337
 Virtuous policy maker, 827  
 Vision zero, 261, 262, 891, 892  
 Vitrified waste, 309  
 Vividness, 683  
 VMPFC. *See* Ventromedial prefrontal cortex  
 Voluntariness, 821, 920, 925, 929  
 Voluntary risks, 252, 253  
 von Neumann, J., 333, 382, 1143  
 von Neumann–Morgenstern representation theorem, 382–387  
*Vorsorgeprinzip*, 966  
 Voting paradoxes, 898  
 Vulnerability, 861, 866–868, 873
- W**
- Wada test, 703  
 Wald, G., 1158  
 Warning labels, 695  
 Warranted choice, 548  
 Waste life-time, 296–298, 308, 309, 313, 314  
 Weak and strong interpretations, 970, 971  
 Weber, M., 1146, 1147  
 Weber’s law, 684  
 Weighted utility, 507–508  
 Weight of the information, 527  
 Weights of attributes, 521  
 Weinberg, A., 1151  
 Welfare-based ethics, 333  
 Welfare economics, 49  
 Well-being, 324–326, 333, 334, 380, 981–982, 984, 987, 990, 995  
 Wernicke’s area, 702–704  
 White male effect, 741, 742  
 WHO, 214, 216, 223, 226, 227, 230  
 Wide reflective equilibrium, 903  
 Wildavsky, A., 1150, 1155  
 Wilde, G., 250  
 Williams, B., 835  
 Willingness
  - to accept, 799
  - to pay, 799, 801, 947–949
  - of trust, 860
 Windelband, W., 1147  
 Wingspread statement, 334, 962, 968, 970–972  
 Wittgenstein, L., 330  
 Wolff, J., 981, 985, 993, 995  
 World’s intrinsic utility, 533–534  
 World Trade Organization (WTO), 1124, 1131, 1133  
 Worldviews, 680, 683, 726–740, 743, 752, 753, 755, 756  
 Worry, 667, 669, 825, 826  
 Worst case, 767, 775  
 Worst-case scenario, 333, 335, 336  
 WRR, 1106–1108  
 WTO. *See* World Trade Organization
- Y**
- Yucca Mountain repository, 306, 308, 316

