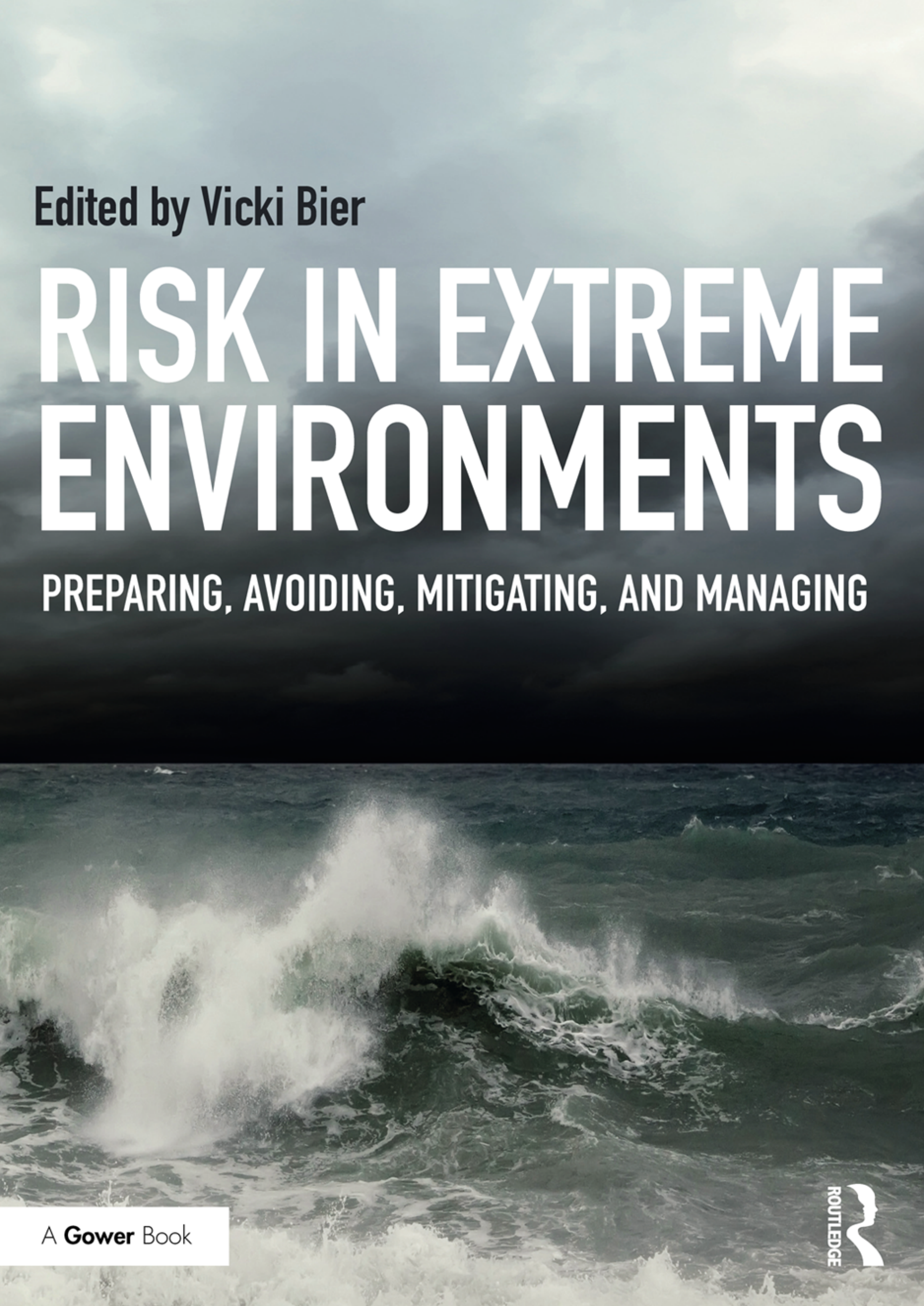
Edited by Vicki Bier

RISK IN EXTREME ENVIRONMENTS

PREPARING, AVOIDING, MITIGATING, AND MANAGING



A Gower Book

ROUTLEDGE

Risk in Extreme Environments

Risk in Extreme Environments presents a wide-ranging discussion of approaches for assessing and managing extreme risks. Extreme events are not only severe, but also outside the normal range of experience of the system in question, and can include environmental catastrophe; engineering failure; and nuclear or other extreme terrorism. The book focuses on synthesizing research results in a way that provides insights useful to decision makers, and enables them to ask probing questions about the risks faced by their organizations, identify creative solutions, and minimize the neglect of extreme risks that can come from an excessive focus on mundane or ordinary management challenges.

The book includes case studies on nuclear power, infectious diseases, and global catastrophic risks. The chapter authors include experts in economics, engineering, geography, law, political science, psychology, sociology, and science in addition to risk analysis.

Risk in Extreme Environments is an accessible and valuable resource for risk managers and other decision makers responsible for complex business and government decisions, while also providing enough detail and references to be informative for risk analysts interested in learning more about technical aspects of the various methods.

Vicki Bier, PhD, is a Professor in the Department of Industrial and Systems Engineering at the University of Wisconsin-Madison. She has written a number of papers on extreme risk, and is a coauthor or editor of four books in addition to this one.

"Risk, extremes, and the steps toward synthesizing and managing them are fundamental challenges that society faces. This book captures these important themes comprehensively, effectively framing them for decision-makers. Bier draws upon her longstanding reputation in these fields and has assembled an exemplary set of international experts from multiple sectors and disciplines to address these critical areas. Together these contributions act as a foundation for advancing risk-based dimensions of decision-making that uniquely draw together diverse perspectives. The triangular relationships among science, organizations and public perceptions are evident in this work. Extending beyond theory, these elements are applied to a rich and enduring set of cases of extremes that reflect the range of decisions that risk managers face."—**Rae Zimmerman, PhD, Professor of Planning and Public Administration, New York University—Wagner Graduate School of Public Service**

Risk in Extreme Environments

Preparing, Avoiding, Mitigating, and Managing

Edited by Vicki Bier

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2018
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
711 Third Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
Names: Bier, Vicki, editor.
Title: Risk in extreme environments : preparing, avoiding, mitigating and managing / edited by Vicki Bier.
Description: Abingdon, Oxon ; New York, NY : Routledge, 2018. | Includes bibliographical references and index.
Identifiers: LCCN 2017023725 (print) | LCCN 2017037780 (ebook) | ISBN 9781315557540 (eBook) | ISBN 9781472439901 (hardback : alk. paper)
Subjects: LCSH: Risk management. | Disasters. | Accidents--Prevention. | Emergency management. | Risk--Sociological aspects.
Classification: LCC HD61 (ebook) | LCC HD61 .R56875 2018 (print) | DDC 658.15/5--dc23
LC record available at https://lccn.loc.gov/2017023725

ISBN: 978-1-472-43990-1 (hbk)
ISBN: 978-1-315-55754-0 (ebk)

Typeset in Bembo
by Saxon Graphics Ltd, Derby

Contents

Illustrations

Figures

Tables

About the editor

Vicki Bier, PhD, holds a joint appointment as Professor in the Department of Industrial and Systems Engineering and the Department of Engineering Physics at the University of Wisconsin-Madison, where she served as chair of the Department of Industrial and Systems Engineering for more than five years. She has directed the Center for Human Performance and Risk Analysis (formerly the Center for Human Performance in Complex Systems) since 1995. She has more than 30 years of experience in risk analysis for the nuclear power, chemical, petrochemical, and aerospace industries, as well as homeland security and critical infrastructure protection. Before joining academia, she spent seven years as a consultant at Pickard, Lowe and Garrick, Inc. While there, her clients included the U.S. Nuclear Regulatory Commission, the U.S. Department of Energy, and a number of nuclear utilities, and she prepared testimony for Atomic Safety and Licensing Board hearings on the safety of the Indian Point nuclear power plants.

Dr. Bier's current research focuses on applications of risk analysis and related methods to problems of security and critical infrastructure protection. Her research has been supported by the U.S. Department of Homeland Security, the National Science Foundation, the Nuclear Regulatory Commission, the Idaho National Laboratory, the Midwest Regional University Transportation Center, the Wisconsin Department of Health and Family Services, and the Wisconsin Office of Justice Assistance. She has also been a special-term appointee for the Risk and Infrastructure Science Center at Argonne National Laboratory, and a special government employee for the Environmental Protection Agency.

Dr. Bier received the Women's Achievement Award from the American Nuclear Society in 1993, and was elected a Fellow of the Society for Risk Analysis in 1996, from which she received the Distinguished Achievement Award in 2007. She is a past president of the Decision Analysis Society, from which she received the Frank P. Ramsey Medal in 2016. She has written a number of papers and book chapters related to the risk of extreme events, and is a coauthor or editor of four books in addition to this one. She served as engineering editor for *Risk Analysis* from 1997 through 2001, and has served as a councilor of the Society for Risk Analysis. Dr. Bier has also served as a member of both the Radiation Advisory Committee and the Homeland Security Advisory Committee of the Environmental Protection Agency's Science Advisory Board.

Contributors

Terje Aven has been Professor in Risk Analysis and Risk Management at the University of Stavanger, Norway, since 1992. He is a Principal Researcher at the International Research Institute of Stavanger (IRIS). He is Editor-in-Chief of *Journal of Risk and Reliability*. He is an Associated Editor of *Journal of Applied Probability on Reliability Theory*, an Area Editor of *Risk Analysis* on Policy and Europe Editor of the *International Journal of Performability Engineering*. He is currently the Chairman of the European Safety and Reliability Association (ESRA) and President-Elect of Society for Risk Analysis (SRA). He was a Director of Gassco AS 2013–2016. Previously he was also Adjunct Professor in Risk and Reliability Analysis at the University of Oslo and the Norwegian University of Science and Technology. He has many years of experience as a risk analyst and consultant in industry. He is the author of many books and papers in the field, covering both fundamental issues as well as practical risk analysis methods. He has led several large research projects in the area of risk, with strong international participation.

Anthony M. Barrett is Director of Research of the Global Catastrophic Risk Institute, which he cofounded in 2011. Barrett is also Senior Risk Analyst with ABS Consulting in Arlington, Virginia. His research interests include risk analysis, risk management, and public policies in a wide variety of areas, including nuclear warfare, emerging technologies, natural hazards, and terrorism. Barrett holds a PhD in Engineering and Public Policy from Carnegie Mellon University and a B.S. in Chemical Engineering from the University of California, San Diego. He was also an ASME/AAAS fellow to the U.S. Department of Homeland Security (DHS), a postdoctoral research associate at the Homeland Security Center for Risk & Economic Analysis of Terrorism Events (CREATE) at the University of Southern California, and a Stanton Nuclear Security Fellow at the RAND Corporation in Arlington, Virginia.

Seth D. Baum is Executive Director of the Global Catastrophic Risk Institute, a nonprofit think tank that Baum cofounded in 2011. His research focuses on risk and policy analysis of catastrophes, such as global warming, nuclear war, and future artificial intelligence. Baum received an M.S. in Electrical Engineering from Northeastern University and a PhD in Geography from Pennsylvania State University. He then completed a postdoctoral fellowship with the Columbia University Center for Research on Environmental Decisions. In addition to his scholarly work, he writes frequently for popular media and is a featured columnist for the *Bulletin of the Atomic Scientists*.

Michael Corradini is Wisconsin Distinguished Professor of Nuclear Engineering and Engineering Physics at the University of Wisconsin-Madison. He has published widely in areas related to transport phenomena in multiphase systems including jet spray dynamics and vapor explosion phenomena. In 1998, he was elected to the National Academy of Engineering. He has also served as a presidential appointee in 2002 and 2003 as the chairman of the Nuclear Waste Technical Review Board (a separate government agency). From 2004 to 2008, he served as a board member of the INPO National Accreditation Board for Nuclear Training. In 2006, he was appointed to the NRC Advisory Committee on Reactor Safeguards and was elected to the National Council on Radiation Protection and Measurements. Most recently, he was appointed Chair of the Scientific Advisory Committee to the French Atomic Energy Agency. He was the first Director of the Wisconsin Energy Institute. He was elected as the President of the American Nuclear Society for 2012–2013.

Louis Anthony Cox Jr., PhD, is President of Cox Associates (www.cox-associates.com), a Denver-based applied research company specializing in quantitative health risk analysis, causal modeling, advanced analytics, and operations research. He holds a PhD in Risk Analysis (1986) and an S.M. in Operations Research (1985), both from M.I.T; an A.B. from Harvard University (1978); and is a graduate of the Stanford Executive Program (1993). He is a member of the National Academy of Engineering (NAE); a Clinical Professor of Biostatistics and Informatics at the University of Colorado Health Sciences Center; Editor-in-Chief of *Risk Analysis: An International Journal*; an Edelman Laureate of the Institute for Operations Research and Management Science (INFORMS); a Fellow of the Society for Risk Analysis (SRA); and a recipient of SRA's Outstanding Risk Practitioner Award. He has taught many graduate and professional courses in risk analysis, decision analysis, and advanced analytics. He has authored and coauthored over 200 journal articles and book chapters on these fields. His most recent books are *Breakthroughs in Decision Science and Risk Analysis* (Wiley, 2015), *Improving Risk Analysis* (Springer, 2013), and *Risk Analysis of Complex and Uncertain Systems* (Springer, 2009).

Aklesso Egbendewe-Mondzozo holds a PhD in Agricultural Economics and a Master of Science degree in Economics from Texas A&M University (USA). He worked first as a visiting assistant professor at Michigan State University. Subsequently, he served as a senior researcher both at the Euro-Mediterranean Center for Climate Change and Fondazione Eni Enrico Mattei in Italy. He is currently an associate professor of Economics at the Université de Lomé (Togo).

Levan Elbakidze is an assistant professor in the Division of Resource Economics and Management, Davis College of Agriculture, Natural Resources and Design, at West Virginia University. He earned his M.S. from the University of Nevada, Reno, and PhD from Texas A&M University. His research program includes management of invasive species, pests, and infectious animal diseases; water resource economics; energy economics; and terrorism. He is also interested in applications of experimental economics particularly for estimation of non-market values. He has received research grants from the U.S. Department of Agriculture, National Oceanic and Atmospheric Administration, U.S. Geologic Survey, and the Nature Conservancy. His work has been published in *Risk Analysis*, *Water Resources Research*, *Ecological Economics*, *Agricultural Economics*, *American Journal of Agricultural Economics*, *International Journal of Risk Assessment and Management*, and other peer reviewed journals.

Ido Erev is a professor of Behavioral Sciences and Economics at the Technion and Warwick Business School. His research focuses on the effect of experience on choice behavior in an attempt to clarify the conditions under which wise incentive systems can solve behavioral and social problems.

Eyal Ert is a professor in the department of Environmental Economics and Management at the Hebrew University of Jerusalem. He studies behavioral processes of decision making under risk and uncertainty, and their personal, social, and environmental implications.

Robert Goble is Research Professor in Environmental Science and Policy and Adjunct Professor of Physics at Clark University. Professor Goble received a B.A. with honors in Physics from Swarthmore College and a PhD in Physics from the University of Wisconsin. After work in theoretical high-energy particle physics at Yale University, the Universities of Minnesota and Utah, and Montana State University, he joined the faculty of Clark University in 1976 with new interests in technology assessment, risks, and hazards. Professor Goble's current research focuses on taking a long-term perspective on environmental and health issues with close attention to uncertainty. His concerns include problems posed by systems of risks, risk management implications of uncertainty, and uses of risk perspectives in international development contexts. Specific efforts address differences among people in their susceptibility to risks from toxics, the nature of regional sustainability, and long-term management of the AIDS epidemic. He has worked previously on energy systems, nuclear accidents, radioactive waste, and models for short-range and regional air pollution. He has also worked on environmental justice projects with Native American communities and with urban neighborhoods. Professor Goble has served on a variety of international, national, community, and project-specific scientific advisory boards.

Franck Guarnieri is Professor at MINES ParisTech/PSL Research University. He is the Chair of the Centre for Research into Risks and Crises (CRC). He leads research on industrial and nuclear safety. He focuses on "engineering thinking" and "extreme situations." He is currently involved in a two-year research project with University of California Berkeley (UC Berkeley) on "Nuclear Safety: From accident mitigation to resilient society facing extreme situations." He is also a designated expert in the French National Research Agency (ANR) and in the European Horizon 2020 program. In 2014, he received the prestigious René-Joseph Laufer Prize from the Académie des Sciences Morales et Politiques of the Institut de France.

Amy Hagerman is an agricultural economist with the United States Department of Agriculture's Animal and Plant Health Inspection Service Veterinary Services in the Center for Epidemiology and Animal Health (USDA-APHIS-VS-STAS-CEAH) on the Epidemiologic and Economic Modeling Team. Dr. Hagerman completed a Bachelor of Science (2004) in Agricultural Economics at Oklahoma State University, and a Master of Science (2005) and a Doctorate of Philosophy (2009) in Agricultural Economics at Texas A&M University. After finishing her doctorate, she completed a postdoctoral program working at Texas A&M University on projects funded by the Foreign Animal and Zoonotic Disease Defense Center, a Department of Homeland Security Center of Excellence. Dr. Hagerman's research is primarily focused on animal health economics, namely threats to U.S. livestock production and global animal product trade from foreign animal disease. This includes elements of public policy examination, applied risk analysis, and decision making, and extending the understanding of estimation methods for foreign animal disease mitigation costs and resource requirements. Prior to joining APHIS in August 2013, she worked for USDA Economic Research Service in the Animal Products and Costs of Production Branch as a research economist.

Roger E. Kasperson is Research Professor and Distinguished Scientist at Clark University. Dr. Kasperson is a Fellow of the American Association for the Advancement of Science and the Society for Risk Analysis. He is an elected member of the National Academy of Sciences and the American Academy of Arts and Sciences. He has authored or coedited 24 books and monographs and 160 articles or chapters in scholarly books and journals. From 1994 to 1998 he was Provost at Clark University and from 2000 to 2004 he was Executive Director of the Stockholm Environment Institute. He is currently preparing a book entitled *Risk Conundrums: Solving Unsolvable Problems*.

Christophe Martin, PhD, is researcher at MINES ParisTech/PSL Research University. Since 2001 Christophe Martin has been involved in various R&D projects with major French companies and European and French institutions. He has conducted research into risk management implementation methods and occupational health and safety risk assessment in several industrial sectors. His specialty is small and medium-sized enterprises, specifically subcontractor relationships. He has also worked with the gas and nuclear industries in identifying organizational failures and designing bespoke tools tailored to these organizations. Since 2014, he has been "Resilience and Safety" Chair hosted by MINES ParisTech and founded by four major private French companies (AFNOR Group, GDF Suez, Total, and SNCF). He graduated from the Political Studies Institute and also obtained a Master of Law from Aix-Marseille University.

Bruce McCarl, PhD, studies the ways that risks facing decision makers influence vulnerability and adaptation decisions in many settings, including bioenergy, climate change, water, and animal disease prevention and reaction. In the animal disease biosecurity field, he has examined disease defense protocol design, trading off investments in prevention and detection with anticipated expenditures for response and recovery when events occur. He also has been involved in a carcass disposal study where attention is devoted to the way elements of the carcass disposal problem influence disease management strategies. He is a fellow or Lifetime Achievement Awardee of the American, Southern and Western Agricultural Economics Associations and was a participant in the Intergovenmental Panel on Climate Change, sharing in the 2007 Nobel Peace Prize. He formerly was lead economist in the DHS funded National Center for Foreign Animal and Zoonotic Disease Defense.

Terrence M. O'Sullivan, PhD, is Director of the Center for Emergency Management and Homeland Security Policy Research, and Associate Professor of Political Science, and formerly a research associate at the Center for Risk and Economic Analysis of Terrorism Events (CREATE) at the University of Southern California—the first Department of Homeland Security Center of Excellence. His research and publications are broadly security risk-oriented, including domestic U.S. and international risk analysis of terrorism, natural disasters (including public health/infectious disease and WMD/biosecurity events), public health law, and environmental security/climate change. He was a 2013 recipient of a National Science Foundation RAPID research award to study citizen disaster response to Hurricane Sandy.

Sébastien Travadel is Senior Lecturer at MINES ParisTech/PSL Research University. Sébastien Travadel's research program is focused on engineering thinking in response to overwhelming risks. Through a cross-disciplinary approach, his work explores the anthropological scope of engineering responses to risks and collective technical action in the face of extreme situations. Sébastien Travadel is a chief engineer for the French Ministry of Transport and Equipment. He also holds a PhD in law and is qualified as a senior mathematics teacher. He has held several managerial positions in the French agency in charge of aircraft accident and incident investigations, notably as Head of Major Investigations. He headed the Airspace and Air Navigation Regulation Department at the French Civil Aviation Authority before cofounding a company that produces decision-support software based on big-data computation.

Ron Westrum is Emeritus Professor of Sociology at Eastern Michigan University. His specialties include the sociology of science and technology, creativity, and the culture of organizations. He has often presented at national and international meetings on systems safety. In addition to shorter writings, he has published three books, the last being *Sidewinder: Creative Missile Design at China Lake*. He is currently working on a book entitled *Information Flow and Organizational Culture*.

Introduction

Vicki Bier

Background and motivation

This book presents a wide-ranging discussion of approaches for assessing and managing extreme risks. The book is intended to be accessible and valuable to risk managers and other decision-makers responsible for large, complex business and government decisions, while still providing enough detail and references to be informative for risk analysts interested in learning more about technical aspects of the various methods.

The authors of the chapters presented here include experts in economics, engineering, geography, law, political science, psychology, sociology, and science, in addition to risk analysis, since effective risk management benefits from an interdisciplinary perspective. Why such a broad range of expertise? Time has shown that effective risk management is inherently multidisciplinary. Early efforts in "reliability analysis," which attempted to make the field of risk management more or less a subspecialty of applied mathematics, were not of much use in practice, since many extreme events turned out to be not random combinations of equipment failures, but rather complicated "comedies of error," caused as much by the foibles of human judgment or poorly designed incentive systems as by weaknesses of engineering design. Being broadly read can prepare decision-makers to ask probing questions about the risks faced by an organization, identify creative solutions from other fields, and minimize the tunnel vision or neglect of extreme risks that can come from being overly focused on mundane or ordinary management challenges. While the authors represented in this book include both researchers and practitioners, the focus throughout is to synthesize research results in a way that will provide insights useful to decision-makers, since risk management is a field where theory and practice go hand in hand.

Many past books have focused on one or another aspect of risk analysis and management—for example, mathematical methods for estimating the likelihood of extreme events, or social-science research on how organizations can best prevent, prepare for, and respond to such events. By contrast, this book attempts to cover multiple aspects of the problem. While this necessarily makes the book less than comprehensive in any one area, the goal is to create cross-fertilization. For example, readers who come to the book looking for tips on organizational approaches to risk management will find some of that, to be sure, but may discover methods of risk analysis that they hadn't been aware of, or case studies that shed new light on problems that they thought they understood. Similarly, readers who come from a mathematical or statistical approach to risk management may find that some of the problems they are trying to solve could perhaps be better addressed at an organizational rather than a mathematical level.

In general, there is rarely if ever a single "right" way of making decisions about extreme risks. The best decision may depend on the decision maker's subjective interpretation of highly incomplete data or imperfect scientific understanding, on an organization's goals (e.g., whether it wants to be a low-cost provider or a highly reliable one), on how much the organization can afford to lose (in the case of financial risks), etc. However, history has revealed many "wrong" decisions about risk, in which poor judgment and/or insufficient foresight led to catastrophic events that could have been prevented. The goal of risk assessment and management as a field is to give decision-makers tools to help minimize such catastrophes, whether they are the result of equipment failures, natural disasters, human errors, or just poorly designed management systems. Of course, the challenges are significant, and no one approach to risk management can eliminate all disasters, but a better understanding of risk management (and of what goes wrong when it is absent or misguided) can certainly help.

The remainder of this chapter presents an introduction to the nature of extreme risk, and an explanation of why it is important, as well as an overview of the rest of the book. Other chapters deal with identification and estimation of risks (both for risks that are reasonably well understand and for unanticipated "black swans"), effective decision-making about risk (especially in complex environments), individual and societal perceptions of risk, and case studies illustrating how the methods and ideas in this volume can be applied in practice.

What is an extreme event?

As pointed out in Bier et al. (1999), extreme events are typically considered those that are extreme in *severity*. Events that are merely unusual or low in *frequency* do not require much attention from risk managers if they are not severe. Similarly, events that are extremely *favorable* also are typically not considered problematic—although in principle they can be (consider for example the challenges of a low-income household suddenly learning how to manage a large inheritance).

However, unusually severe events occur only *rarely*. Adverse events that occur frequently, and are within the normal range of experience, are typically ones for which society has already evolved coping mechanisms—although that may vary from one organization to another (for example, active-shooter situations may be considered relatively common in the military, but not in elementary schools). Therefore, throughout most of this book, we will focus on events that are both rare and severe.

Note that such extreme events pose several types of problems. First, the severity of their consequences may make it difficult for an organization to cope with such events when they occur. Their rarity can also make it challenging for organizations to learn about such events—whether from their own experience, from data, or from published analyses. Moreover, people, engineered systems, and organizations may behave differently in extreme events than in normal circumstances, and that behavior may not always be well understood. Finally, even when extreme events are not characterized by high levels of scientific uncertainty, they may nonetheless be difficult for people to grasp and put into perspective, due to their severity and unfamiliarity.

Clearly, no one book can hope to address all of the above problems in detail, or do justice to the entire field of extreme risks. In fact, there are numerous topics that are either omitted from this book, or touched on only in passing. Those include some methods of risk analysis (such as the use of catastrophe theory and chaos theory in finance, or the use of Monte Carlo simulation, extreme-value theory, and/or expert judgment to help estimate risks for which there is little empirical data), some approaches to risk management (such as the use of catastrophe bonds in the insurance field, or reliance on the precautionary principle), some practical methods for making decisions about risks (such as business-continuity planning and enterprise-risk management), and numerous examples of extreme risks beyond those discussed in this book (including climate change, flooding and dam failure, species extinction, systemic financial risk, and terrorism, to name just a few). However, the perspectives on the field provided in this book should provide insights into how to address many challenges involved in managing extreme risks, inspire decision-makers to think about how their organizations could do a better job of preparing for and managing extreme events, and provide references to the literature for readers who want to learn more about specific methods or approaches.

Overview of this book

This book is divided into four parts:

- method for assessing the risks of extreme events
- methods for managing the risks of extreme events
- perceptions of extreme events
- sample applications and case studies.

With regard to *risk assessment*, Cox and Bier describe probabilistic risk analysis and related methods, with several examples. They also briefly discuss how to communicate the results of probabilistic risk analyses to decision-makers, and how to use analysis results to improve decision-making. Although Cox and Bier do briefly discuss methods of hazard identification, the method of probabilistic risk analysis that they discuss is fundamentally limited to risks that can reasonably be identified and quantified. By contrast, Aven discusses the treatment of "black swans"—surprises that may not be at all straightforward to anticipate and quantify before they occur, but are nonetheless a crucial aspect of risk management.

With regard to *risk management*, Westrum describes what the field of organizational sociology has learned about how to achieve excellent performance in extreme situations. He provides numerous historical examples, some in which excellent performance was achieved and some where it was not. Westrum also highlights the importance of information flow as a key resource in achieving excellent performance—not only making sure that people have the information they need to do their jobs, but also ensuring that the lessons of history are embedded in organizational rules and practices. Travadel et al. then describe performance in a truly extreme environment—the nuclear core-melt accident in Japan in 2011. They note that making decisions in such an extreme situation departs from standard models of decision-making, in the sense that responding to the Japanese disaster required not only good information processing, but also leadership, heroism, trust in one's instinctive reactions, and a sense of collective (rather than just individual) responsibility and action—traits that might ordinarily be considered more typical of the military than of industrial organizations.

Returning to the less dramatic task of emergency planning before a disaster occurs, Hagerman et al. discuss the important question of how much an organization should invest in prevention and preparedness for possible disasters, versus in emergency response and recovery. They note that prevention and preparedness will tend to be more worthwhile for events that occur more often, events whose potential consequences are more severe, and events that develop extremely rapidly (for which it may not be possible to implement response and recovery actions quickly enough to keep up with rapidly worsening consequences). Although Hagerman et al. focus their discussion on the case of a possible livestock disease, much the same principles would apply in deciding how much to invest in prevention versus response for a pandemic that affected humans (such as avian influenza), and even for other impending threats (such as climate change).

Goble focuses primarily on the importance of adapting our prevention and response strategies to changing situations. This is important for two reasons: first, because even an anticipated emergency is unlikely to be an exact match for previous preparedness efforts; and secondly, because preparedness can also enhance recovery from unanticipated black swans. Moreover, since extreme events will typically be subject to great uncertainty, new information may become available even before an emergency transpires, again requiring vigilance and subsequent adaptation. However, Goble notes (like Westrum) that adaptive management in the face of extreme events poses significant challenges (with the result that many organizations do not successfully learn from past experience), and provides suggestions for how to improve organizational capabilities for adaptive management.

With regard to *risk perceptions*, this book discusses both individual and societal perceptions. At the individual level, Ert and Erev discuss the results of a number of psychology experiments explaining why people often underestimate risks (e.g., deciding that it's safe to walk in a dangerous neighborhood, merely because they have walked there without incident a few times), but sometimes overestimate risks (e.g., declining beneficial medical treatments because of a small risk of side effects). Over the past decade or so, psychology research has clarified that underestimation of risks is more common when people estimate risk directly from their own experience (e.g., walking in a dangerous neighborhood), while overestimation is more common when people are provided with descriptions of the risk (e.g., risk analysis results, newspaper articles, or brochures)—perhaps because the mere availability of a description makes the risk seem prominent. Those readers not interested in the detailed experiments by which this pattern of behavior has been explored and understood may wish to focus on the section entitled "The impact of extreme risks and potential applications" in the chapter by Ert and Erev, where they discuss the implications of this type of research for risk management. For example, they note that the tendency to underestimate risks based on personal experience can result in decreased use of safety features over time, insensitivity to warnings, and the consequent necessity for continuous enforcement of safety regulations (such as safe workplace practices).

By contrast, Kasperson emphasizes political and societal perception of risks, shedding light on the question of why some extreme risks are tolerated by society with little public concern (due to "risk attenuation"), while other comparable hazard events may elicit much stronger public reactions (a phenomenon known as "risk amplification"). He discusses several recent examples, and notes that failure to account for social processes associated with risk amplification can result in unanticipated adverse consequences such as loss of market share, higher insurance costs, political demands for increased risk regulation, loss of trust, and even protests or other expressions of public outrage.

Finally, *case studies* are provided to illustrate how the ideas in this book can influence actual decision-making. In particular, Corradini and Bier discuss how the nuclear power industry has used probabilistic risk analysis, scientific experiments, and learning from past experiences to manage the risks of a high-hazard technology. O'Sullivan discusses the challenges posed by pandemics and other infectious diseases, and offers practical steps that even organizations outside the medical field can take to prepare for and respond to those challenges. In a more speculative vein, Baum and Barrett discuss how the methods and ideas of risk assessment could be used to analyze risks that can threaten human civilization as a whole (such as asteroid impact or extreme climate change). Although obviously challenging to analyze, global catastrophic risks have arguably received less attention than would be merited by the severity of their consequences, emphasizing the need for research.

This book is intended to stimulate risk managers, policy makers, and other decision-makers to think more deeply about the risks that their organizations face and how to manage extreme risks more effectively. Every organization faces its own set of challenges, so readers will undoubtedly find some chapters to be more relevant to their situations than others. For example, managers who are involved primarily in budget setting may be especially interested in the work of Hagerman et al. on the relative importance of investment in emergency preparedness versus response. By contrast, those who manage first responders may find the lessons of Travadel et al. and Westrum to be directly relevant to the challenges they face. Likewise, decision-makers concerned primarily with managing known risks might be interested in the methods discussed by Cox and Bier, while those more concerned with anticipating unknown risks might find the work of Aven (on black swans) and Goble (on adaptation) to be more relevant.

However, I hope that readers will take the time to explore chapters that may not seem to be directly relevant to their situations. For example, the observations of Travadel et al. about the challenges faced by plant staff after the Fukushima disaster raise questions not only about nuclear safety or about how to manage emergency responders, but also about the nature of leadership in crisis situations; those broader lessons may be thought-provoking for leaders from entirely different fields, such as business or public health. Similarly, the chapters by Goble and Westrum cast doubt on whether bureaucratic plans for emergency response leave organizations better prepared for all types of disasters, or whether the mere act of formulating detailed plans can cause organizations to focus too much on some types of risks and not enough on others. Finally, Ert and Erev provide insights from the field of psychology about why people often act complacently about extreme risks, under the assumption that rare events "won't happen to me," and discuss the implications of this phenomenon for a variety of risks, from theft to terrorism. They also suggest (based on experimental evidence) that active enforcement of rules about safety practices may be more effective than simply providing warnings.

In reviewing some of the key ideas that emerge from this book, two seemingly contradictory themes emerge. One is that methods already exist to help deal with many of the extreme risks facing modern society. For example, Corradini and Bier describe how the methods of probabilistic risk assessment outlined by Cox and Bier have been applied to make nuclear power safer in the U.S. They indicate that even the dramatic events that occurred at Fukushima were not unanticipated, and perhaps could have been prevented by "a coherent risk-informed regulatory approach, utilizing the most advanced evaluation methodologies"; they also note that the U.S. Nuclear Regulatory Commission and the nuclear industry used the Fukushima experience as a basis for identifying strategies to deal with similar extreme events should they occur in the U.S. At the same time, the chapters by Aven (on black swans) and Goble (on adaptive strategies for risk management) suggest that existing methods are at best imperfect and incomplete, and speak to the need for vigilance to identify signals of problems (possibly even before they become truly extreme) and adapt accordingly. Westrum similarly notes that management in extreme environments poses significant challenges, and that not all organizations achieve good performance. This discrepancy (between the availability of useful methods and the continued existence of unsolved problems) is not necessarily a contradiction, but a sign that risk management is an emerging discipline that still retains aspects of an art form—as Travadel et al. suggest when discussing the need for emergency responders at Fukushima to go "beyond any procedures."

Based on the breadth of perspectives in this book, I believe that almost every risk analyst, risk manager, or decision maker will find something of value, whether new methods, perspectives complementary to their own discipline, or a broader understanding of existing controversies and debates. I also hope the ideas in this book will inspire new dialogue between industry and government practitioners, policy makers, and risk analysts.

Reference

Bier, V. M., Y. Y. Haimes, J. H. Lambert, N. C. Matalas, and R. Zimmerman (1999). "Assessing and managing the risk of extremes." *Risk Analysis*, 19, 83–94.

Part 1

Assessing the risks of extreme events

1 Probabilistic risk analysis

Louis Anthony Cox Jr. and Vicki Bier

How well can contemporary organizations and societies design, build, and operate complex engineering systems safely and reliably for long periods? Being able to do so is crucial if nuclear power is to be a viable option, if infrastructure (such as advanced transportation systems and energy distribution networks) is to be trustworthy, if minerals and petroleum are to be discovered and extracted safely, and if hazardous manufacturing and chemical storage facilities are to be located in convenient proximity to transportation hubs and population centers. This chapter, which is an update and extension of Bier and Cox (2007), discusses methods for quantifying the extent to which complex engineering systems can be designed and operated safely. It introduces methods of probabilistic risk assessment (PRA), sometimes called probabilistic safety assessment (PSA), and briefly discusses its potential, some limitations, and various areas of application. The opening section describes and illustrates key concepts of risk analysis and PRA, and explains how it can be used to inform and improve risk management decisions. The second section introduces core techniques of PRA modeling: hazard identification; fault trees and event trees; and frequency-severity diagrams ("F-N curves"). The next two sections consider how the inputs to these and other risk models can be obtained, using both subjective estimates and alternatives such as accident precursor analysis; and discuss techniques for characterizing and bounding uncertainty in risk results, given uncertainties about the inputs and about the validity of the risk models used to calculate outputs (such as accident probabilities) from the inputs. We then turn to the key topic of how to communicate the results of a PRA, before diving more deeply into how to use PRA results to improve risk management decisions. A special feature of this section is its discussion of game-theoretic methods for improving practical risk management decisions about terrorism and other areas whether risks and outcomes depend on the interactions of multiple decision-makers. The final section provides closing thoughts on the use and potential of PRA to make systems safer.

Opinions about whether complex social-economic-environmental-engineering systems can ever be designed and operated safely over long periods are divided. One school of thought, sometimes called *normal accident theory* after the book that articulated it (Perrow 1984), holds that engineered systems with high "interactive complexity" and "tight coupling" of interdependent components or subsystems (so that changes propagate quickly among them) are inherently unpredictable and uncontrollable by human operators. Resulting accidents and catastrophic failures in such high-risk technological systems are seen as inevitable and unavoidable; in this sense, they are "normal." In this pessimistic view, adding redundancy to complex systems to reduce accidents makes them even more complex and prone to unpredictable failures. Case studies of accidents and near-accidents at chemical plants, nuclear reactors, airports, and other complex industrial facilities have been used to illustrate normal accident theory.

A different view, popularized in the catchphrase "Failure is not an option" (made famous by the movie *Apollo 13*), is that complex engineering systems *can* be built and operated safety by sufficiently disciplined, creative, well-organized, and well-trained teams and organizations. Sociologists, psychologists, and other researchers have sought common features of "high-reliability organizations," meaning organizations with significantly fewer accidents and failures than normally expected. They have proposed features such as preoccupation with detecting early signs of failure; reluctance to (over-) simplify interpretations or to prematurely reject unexpected interpretations of observations; sensitivity to operations; commitment to resilience; and appropriate deference to expertise (as opposed to rank, seniority, or power) (Weick and Sutcliffe, 2007). Such habits can help to create the vigilant "mindfulness" needed to operate safely, and to catch and correct potential problems before they cascade out of control. Routinely safe operations on aircraft carriers and in other high-stress, risky, and complex environments vividly illustrate that well-trained teams in well-designed organizations and environments can manage risks successfully. Similar principles can be applied in different settings, such as operating rooms and intensive care units. Eliminating "mindlessness" (e.g., blind rule-following or deference) in the implementation of federally and locally funded programs to reduce infant mortality and pre-term birth has been proposed as a way to reduce the frequency and severity of poor outcomes (Issel and Narasimha, 2007).

Applying PRA to engineered systems gives engineers and risk managers practical tools to understand, predict, and manage risks for a variety of complex engineered systems. PRA identifies how systems might fail, the likely (and not-so-likely) potential adverse consequences of failures, and how best to prevent failures and mitigate adverse consequences while meeting other goals, such as continued productive operation of a hazardous facility. PRA methods include probability modeling techniques (both analytic and simulation-based) for quantifying engineering risks, typically expressed as probabilities of adverse events and as frequencies and severities of their adverse consequences over a stated period of time. PRA also includes optimization methods from operations research and safety and reliability engineering that can identify cost-effective ways to improve safety and reliability while satisfying other constraints (e.g., constraints on system cost, weight, or operating performance).

Introduction to probabilistic risk assessment

PRA is a method that helps analysts and decision-makers to understand and describe risky situations, learn about them from data, discuss them productively, and decide what to do about them. It addresses how to predict the probable consequences of alternative choices (*risk assessment*), including the risk of doing nothing. It is also useful in explaining and presenting risk information to others (*risk communication*) and in informing deliberations and choices among alternative courses of action (*risk management*), even if the information available when a decision must be made or justified is far from perfect. Figure 1.1 illustrates the risk manager's essential problem: deciding what to do next when the consequences of each choice are uncertain.

Figure 1.1 Risk management provides principles for answering "What should we do next?" when the consequences of different choices are uncertain. PRA helps risk managers to identify clear, defensible rationales for their choices and recommendations, rather than simply asserting "Here is what I think (or feel) we should do next."

Formal PRA sometimes yields unexpected or counterintuitive answers about the relative sizes of different risks or the relative benefits of different risk management alternatives. Key factors that do not enter into formal quantitative risk assessment, but nonetheless affect perceptions of risk, include perceived benefits, fairness, and the political or ethical legitimacy of hazardous activities. Conversely, risk perceptions often place little or no weight on the frequencies and severities of potential losses (see, for example, Kahneman et al., 1999; Sunstein, 2002), which are crucial drivers of PRA results. Public and political concerns about perceived risks, and demands for actions to reduce them, often reflect outrage more than actuarial calculation of frequencies and severity distributions.

For high-stakes decisions, a manager might want to use more than intuition, public outrage and demand for action, or informal guess-work about which choice to make or recommend. This is especially true if she must explain or justify a recommendation to others. PRA helps by providing the following components for structuring and justifying decisions.

- *Risk assessment* (Kaplan and Garrick, 1981). This step describes uncertainties about the consequences caused by different choices. In PRA, uncertainties are presented as *probabilities* for the different possible consequences of each choice. PRA develops and documents the data, models, calculations, and assumptions used to quantify these consequence probabilities. In general, PRA identifies possible consequences of different choices and provides some information about how likely they are for each choice, using probabilities (rather than possible alternatives such as qualitative ratings or rankings, linguistic descriptions of uncertainty, or fuzzy interpretations of inherently ambiguous linguistic terms such as "unlikely.")

- *Risk evaluation and comparison* (Montibeller and Franco, 2007). This step quantifies and compares the values of different choices for managing risks. It may require making value trade-offs, such as between the sizes of potential benefits and the probabilities of getting them; between larger potential losses and the chance for larger potential rewards; between smaller rewards sooner and larger ones later; and between present costs to some people and uncertain future benefits to others.
- *Risk communication* (Bier, 2001a, 2001b). PRA outputs can be structured and formatted to help present and explain risk information clearly, vividly, and persuasively, so that others can understand and act on it. Effective risk communication benefits from an understanding of how risk information presented in different ways affects perceptions, beliefs, behaviors, and engagement of different stakeholders in risk management deliberations.
- *Risk management* (Paté-Cornell and Cox, 2014). The previous components are intended to support improved decisions about how to manage risks. Risk management recommends some choices over others. Risk management decision processes not only produce recommendations for what to do, but also include monitoring the implementation and results of decisions, and learning from experience about what works, how well it works, and how to improve future decisions.

Risk assessment, risk communication, and risk management are frequently identified as key components of risk analysis. Outside of financial and economic applications, risk evaluation is not usually identified as a separate component, but is subsumed in the other steps. Risk communication bridges between the scientific world of risk assessment and the policy world of risk management, communicating the results of risk assessments to inform risk management deliberations and choices. Risk communication also addresses how to present facts, data, and policy recommendations in ways that demonstrably change behaviors to reduce risks.

The ultimate goal of PRA is usually to understand and compare the uncertain consequences of different courses of action in ways that can be explained to, understood by, and independently verified by others. Many risk analysis problems involve managing threats to life, health, safety, the environment, peace and security, way of life, interests of possible far-future generations, and other traditionally hard-to-quantify values. Moreover, most important risk management decisions are made by more than one person, perhaps after considerable deliberation and presentation of differing concerns, points of view, and recommendations. Thus, risk analysts should be expert in using other disciplines to help individuals and groups make well-supported and vitally important decisions under uncertainty.

Some examples of risk management decisions that could in principle be analyzed by PRA include choices among the following alternatives:

- medical treatments for a patient admitted to an emergency room, when the underlying diagnosis and the patient's responses to different treatment choices are uncertain;

- financial portfolios, R&D portfolios, or portfolios of risk-reducing activities in which an individual or a corporation might invest;
- emissions levels that a regulatory agency might permit from power plants;
- flood insurance policies that an individual or a business might buy;
- levels of inventory that a retailer chooses to keep on-site, in order to meet demand and protect against possible disruptions in supply;
- security measures at airports;
- whether to pursue a lawsuit or settle out of court;
- whether to walk down a dangerous street in a city at night;
- whether to declare war in response to provocation by another country.

Even personal choices about where to live and work, what career decisions to make, or whether to walk down a dangerous street at night could in principle be analyzed within the framework of risk analysis. However, formal PRA is usually applied only to decisions that affect many interested parties, and/or that require a well-documented and explicit rationale (e.g., high-stakes decisions about medical treatments).

Examples of complex engineering systems to which PRA has been successfully applied include: nuclear power plants, beginning with the Reactor Safety Study (U.S. Nuclear Regulatory Commission, 1975) and continuing to the present day; the space shuttle (both before and especially after the Challenger disaster; Paté-Cornell and Fischbeck, 1993); dam and reservoir planning and operations (Lave and Balvanyos, 1998); highway, bridge, and transportation infrastructure (Leung et al., 2004; McGill et al., 2007); emergency planning (Lv et al., 2013); liquefied natural gas terminals and other flammable gas storage and transportation facilities (Defriend et al., 2008); facilities and operations that might release hazardous substances (Till et al., 2014); and electric power generation and distribution planning (Mili et al., 2004). The common elements in such systems are that they all involve: (1) a *designed system* intended to withstand different levels of stress, with the option of incorporating different levels of backup and fail-safe design; (2) a *system operator/risk manager* faced with decisions about how to inspect, maintain, and use the system (e.g., when to launch, when to shut down, and generally what level of precaution to adopt); and (3) an uncertain *environment* that generates stresses and adverse conditions that the system should ideally be able to withstand. Uncertainties from the environment may involve: random events, such as equipment failures or unexpectedly high or stressful transient loads (as in the case of the Tacoma Narrows bridge collapse; https://en.wikipedia.org/wiki/Tacoma_Narrows_Bridge#Original_bridge); natural disasters (such as earthquakes, floods, or hurricanes); terrorist attacks; or operator errors (perhaps arising from miscommunication, lack of coordination, or misunderstanding of system behavior among those running a system). Unexpected behaviors of interacting software modules or other subsystems can also cause a system to fail, even if each component performs as designed (Leveson, 2004).

PRA is usually applied to rare and catastrophic events for which it may be difficult to estimate risks directly due to lack of empirical data, the possibility of unobserved changes (e.g., deterioration) in the system, and changes in the system's environment or use. Risk assessment can also be applied to predict routine (e.g., occupational accident) risks, although in such cases, it may be possible to rely primarily on empirical data, reducing the need for modeling. In general, PRA is used to estimate, predict, and find ways to reduce the risks to facility or system owners, employees, and the public.

Risk analysis can help to inform design decisions (e.g., trade-offs among safety and performance, cost, etc.) and also operational decisions (e.g., when to shut down a facility). It can be useful regardless of who makes the decisions—for example, facility owners and operators, regulators, or multiple stakeholders interacting through a participatory risk management and conflict-resolution process (as described in Stern and Fineberg, 1996). Key technical challenges that PRA must address include: how to predict the performance and quantify the behaviors of a complex system, given the system's design and the operator's decisions, in the face of inadequate data; how to optimize the joint decisions faced by the system designer and owner/operator (which can involve hard combinatorial optimization problems, as well as problems of coordination and communication between different organizations); how to most effectively model interdependencies and uncertainties about the system's current state; the development of cost-effective "screening" methods for addressing the myriad possible risks in "open" systems (such as the risk of terrorist attack); and the scaling-up of problems for extremely complex systems, such as infrastructure networks. Also, there is still room to benefit more fully from adaptation of methods developed in other fields, including decision analysis and related fields (such as Bayesian statistics). The following sections expand on these points.

Fundamentals of PRA models: hazard identification, event trees, fault trees, *F-N* curves

PRA typically begins by defining a system to be analyzed and identifying undesired outcomes that might occur when it is operated. *Hazard identification* methods have been developed to identify potential adverse consequences of system operation. Structured qualitative techniques include hazard and operability (HAZOP) studies, and failure modes and effects analysis (FMEA), which describes potential failure modes, causes, effects, safeguards, and recommendations for reducing risks.

Fault trees and event trees can be used in a qualitative mode for hazard identification, but can also be quantified to estimate the likelihood of adverse events. *Event tree analysis* starts with an "initiating event" and works forward to identify its potential consequences. It shows potential sequences of events, with the probability of each branch leaving an event node (representing the possible resolution of an uncertainty, often modeled as a possible value of a random variable) being conditionally independent of earlier information, given that the branch point (i.e., that event node) has been reached. The frequency of a given event sequence is then just the product of the conditional branch probabilities along that path, multiplied by the frequency of the initiating event.

Fault tree analysis (Barlow 1998) begins with an undesired outcome, called the "top event," and reasons backward to identify which combinations of more basic events (e.g., component failures) could bring about the top event (e.g., failure of the system). The result is a tree that represents those sets of basic events that would be sufficient to cause the top event using "AND" and "OR" logic (and possibly more complicated logic gates as well). The tree generally goes down to the level of basic events whose probabilities can be reliably estimated from experience, judgment, and/or data. Both fault trees and event trees can be represented as logically equivalent influence diagrams and solved by well-developed influence diagram algorithms (Barlow 1998; Bobbio et al., 2001).

Example: fault tree calculations for car accidents at an intersection

Setting

Suppose that a car accident (the top event) occurs at an intersection if and only if (two cars approach the intersection at the same time from different directions) AND (both cars proceed). The event "both cars proceed" can be further decomposed into a logical sub-tree, as follows: (both cars proceed) if and only if [(the signal is broken AND both cars proceed) OR (the signal is not broken AND both cars proceed)]. Reliable statistics may show that the first event (sometimes called the *initiating event*) – namely, "Two cars approach the intersection at the same time from different directions" – occurs with an average annual frequency of, say, 100 times per year. The signal is broken on 1% of these occasions (independently of traffic), and the conditional probability that both cars will proceed following the initiating event is 0.1 if the signal is broken, and 0.01 if it is not broken.

Problem

(a) What is the average annual frequency of accidents at the intersection, given these numbers? (b) What fraction of accidents would be prevented if the signal were made perfectly reliable, so that it never failed?

Solution

(a) The average annual frequency of accidents is (100 simultaneous approaches per year) × (1% probability that signal is broken when a simultaneous approach event occurs) × (0.1 collision probability if signal is broken) + (100 simultaneous approaches per year) × (99% probability that signal is working when a simultaneous approach event occurs) × (0.01 collision probability if signal is working) = $100 \times (0.01 \times 0.1 + 0.99 \times 0.01) = 1.09$ accidents per year. (b) If the signal were perfectly reliable, the accident rate would be reduced to $100 \times 0.01 = 1$ accident per year. Thus, $0.09/1.09 = 8.26\%$ of accidents would be prevented by eliminating signal failures.

In this example, accidents with the traffic signal working constitute a *dominant contributor* to the average annual accident frequency. This means that ignoring other, rarer events (namely, accidents with the signal broken) yields approximately the same calculated risk number (about one expected accident per year). One way to simplify fault tree calculations is to focus on dominant contributors, neglecting events that are rare enough so that they do not change the numerical answer (within some desired level of precision). If reliable statistics were not available for the relevant probabilities, they might be estimated from experiments (e.g., using driving simulator results), models of driver behavior, or expert judgment. Uncertainty and sensitivity analyses would then typically be used to determine by how much the calculated risk might change if different plausible estimates or better future information about these inputs were to be used in the analysis. Each event in the model, such as "both cars proceed," could potentially be expressed as a sub-tree consisting of a logical combination of more refined event descriptions, e.g., (both cars proceed and weather is good) or (both cars proceed and weather is not good). Refinement of descriptions stops when the current description allows basic event probabilities to be quantified accurately enough to support risk management decisions.

More generally, a *quantitative risk model* typically consists of a mathematical or computational model of a system of interest, together with consequence attributes of interest, and one or more alternative risk management decisions to be evaluated or decision variables to be optimized. The model can be used to predict the probable consequences of alternative decisions, and identify preferred decisions—those that yield preferred probability distributions (or, more generally, preferred stochastic processes) for the consequences of interest. Risk modeling typically involves some or all of the following components.

- *System representation* (Barlow 1998; Smith 2005). An engineered system is often represented mathematically in one of the following forms:
  (a) a "black-box" *statistical model* (e.g., a "hazard function" quantifying the failure rates of a system at different ages or elapsed times, given that it has not failed so far);
  (b) component failure rates combined via fault trees or event trees (or generalizations called coherent structure functions) that determine the system's state (e.g., working or failed) from the states of the components (a coherent structure function must be monotonically increasing, going from a system failure probability of zero if all components work to a system failure probability of one if all components fail; Barlow and Proschan, 1975);
  (c) a stochastic *state transition* model (e.g., a Markov or semi-Markov model for transitions among working and failed components, representing component failure and repair rates; Howard, 1971);
  (d) a discrete-event *simulation model* (Smith 2005).
- *Environment representation*. Like a system model, a model of the environment may be a statistical black-box model (e.g., a function describing the frequency and intensity of stresses to the system's components), a stochastic process, or a simulation model. Plausible worst-case or bounding scenario analyses are sometimes used when probabilistic descriptions of uncertainty are unavailable or difficult to obtain. A model of the environment is often incorporated directly into the system model, as with traffic levels and weather conditions in our hypothetical traffic-accident model.
- *Decision-rule representation*. A *decision rule* for managing an engineered system maps observed information about the system into a resulting action or intervention. For example, a component may be replaced based on the observed history of failures and repairs. Optimization methods, including recently developed simulation-optimization techniques (see for example Ólafsson and Kim 2002), can help to identify "good" or "best" decision rules given a system model, an objective function (e.g., a multiple-attribute utility function), and a model of the environment. Of course, many decisions in the real world (even when informed by PRA) are made without a formal decision rule, either because the PRA results themselves make the best decision clear, or because of the need to address concerns of multiple stakeholders, etc.

Example: bug-counting models of software reliability

To illustrate the idea of a black-box model as discussed above, we here discuss the use of statistical models to track the number of remaining computer bugs in software. In such a "bug-counting" model, the (unknown) initial number of bugs in a piece of code is represented by a random variable N with a prior distribution. As the code is tested and debugged, the remaining number of bugs presumably decreases, and the random times between successive bug discoveries stochastically increase. (Relatively sophisticated models also allow for the possibilities that detection and repair are imperfect processes and that debugging activities may introduce new bugs.) The empirical record of bug discoveries can be used to trigger a decision rule such as "If no bugs have been discovered within M tester hours, then release the software." Simulation optimization can then be used to numerically optimize the parameter M.

Example: risk management decision rules for dams and reservoirs

To illustrate the idea of a decision-rule representation as discussed above, we here present a set of decision rules for managing water releases from dams and reservoirs, as given by Wurbs (2005, p. 18):

> Release decisions depend upon whether or not the flood control storage capacity is exceeded … federal reservoirs are typically sized to contain at least a 50-year recurrence interval…flood and, for many projects, design floods greater than the 100-year flood …, perhaps much greater. A specified set of rules, based on downstream flow rates, are followed as long as sufficient storage capacity is available to handle the flood without having to deal with the water surface rising above the top of flood control pool … For extreme flood events which would exceed the reservoir storage capacity, moderately high damaging discharge rates beginning before the flood control pool is full are considered preferable to waiting until a full reservoir necessitates much higher release rates.

The outputs from quantitative risk models are often summarized as *F-N* curves (also sometimes called exceedance probability curves, or complementary cumulative frequency distributions), showing the expected annual frequency F of fatalities or damages exceeding any given level, N, for $N \geq 0$. See Figure 1.2 below for an example. Other risk displays show how risk varies by location, over time, and with other characteristics. For example, it is common practice to plot "risk contours" showing risks to individuals at different locations around a potentially hazardous installation or transportation route.

Major technical challenges for developing PRA results include:

1. *Constructing and validating models* of the system and its environment. Statistical analysis of accident precursors uses data on "near misses" to validate and refine model-based predictions (Borgonovo et al*.*, 2000; Phimister et al*.*, 2004; Yi and Bier, 1998). For example, data on situations in which an accident "almost" happens can be used to validate whether existing models of accident causation are sufficiently complete to capture all observed causes; thus, the fire at the Browns Ferry nuclear plant made clear that comprehensive risk analyses for nuclear plants needed to include models of fire initiation and propagation. Powerful model-building and model-checking methods have also been developed in the areas of *system identification*, which attempts to identify dynamic system descriptions of input-output relations from observed time course data (Plett, 2003), and *data mining and machine learning*, which seek to learn correct models (or at least subsets of especially plausible models) directly from data (Kourou et al., 2014; Montazeri et al., 2016).
2. *Calculating, simulating, or estimating probabilities of rare events.* Methods for addressing this challenge have advanced significantly in recent years (see for example Bucklew, 2004; Rubinstein and Kroese, 2004).
3. *Treatment of dependencies* among failure events and system components. Methods for treatment of dependencies (Mosleh et al., 1998) presently include common-cause failure analysis (to show dependence in the failure rates of similar components due to a common underlying cause); dependency matrices and event trees (to show dependence of some systems on "support" systems such as electric power); and external-events analysis (to capture the fact that events such as earthquakes, fires, and floods can affect multiple components of a system).

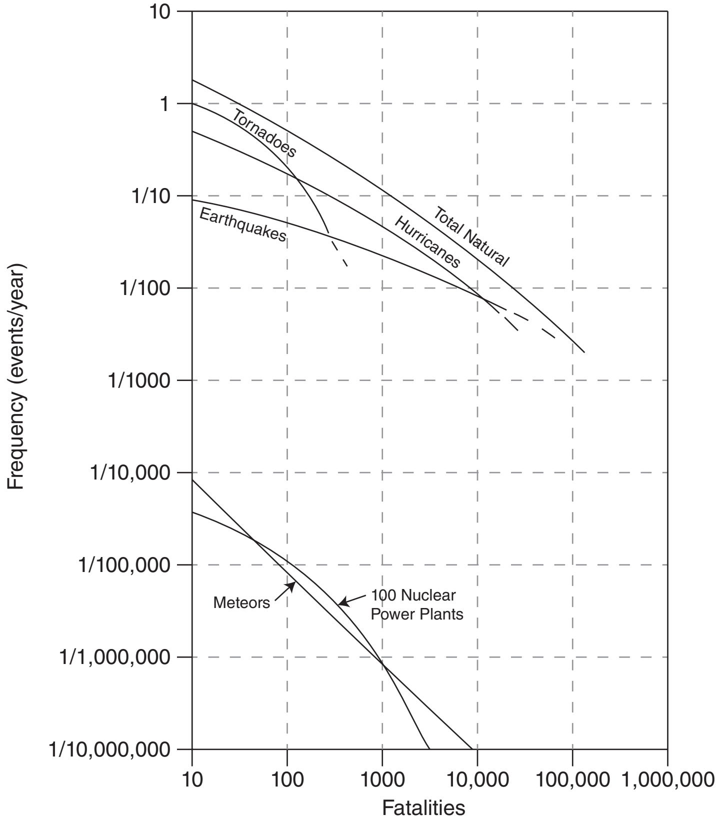


Figure 1.2 Frequency of fatalities due to natural events.

Quantifying model components and inputs

A model typically expresses risk (e.g., probability of failure by a certain time) as a function of the performance of model components and/or input parameters. These must be quantified from available data, perhaps using a combination of expert judgment and Bayesian statistics (due to the sparseness of directly relevant data). In Bayesian statistics, a subjective prior distribution (e.g., based on expert opinion) is updated with observed data to yield a posterior probability distribution for the quantity of interest (Lee 2004). Thus, Bayesian methods balance how likely a particular event was thought to be before evidence was available (based on expert opinion) against how consistent that opinion is with the new evidence.

Although Bayesian approaches to quantifying risk models are frequently applied in practice, advances are still being made in numerous areas. These include: designing more flexible and tractable models for treating probabilistic dependence in risk models; alternatives to relying on subjective prior distributions (which can be problematic if plausible differences in subjective priors significantly affect risk results); and treatment of model uncertainty (e.g., uncertainty about which variables to include in a model, and/or the most appropriate functional form for a model). Each of these is discussed below.

Modeling interdependent inputs and events

If the state of a system is described by a coherent structure function, and each component independently undergoes stochastic transitions over time (e.g., from "working" to "failed" to "repaired" or "replaced"), then the probability distribution for the system's state (i.e., the probability that it will be working rather than failed at any time) can be obtained relatively easily. Stochastic simulation of the behaviors of the components, and routine application of combinatorial reliability models and algorithms (such as fault tree analysis or event tree analysis), are practical even for large systems. However, if component behaviors are interdependent (e.g., if each component failure increases the stress on those components that have not yet failed), then it becomes more complex to calculate the risk that the system will have failed by any given time. Simulating interdependent behaviors may be straightforward in principle, but, in practice, it requires specifying how events depend on each other—a potential combinatorial nightmare.

Dependence can also be a problem for uncertainty analysis. In particular, the failure rates (or probabilities) of the various components can be uncertain and statistically dependent on each other, even if their behaviors are conditionally independent given their failure rates. For example, learning that one component had a higher failure rate than expected may cause one to increase estimates of the failure rates of other similar components. Failure to take such dependence into account can result in substantial underestimation of the uncertainty about the overall system failure rate (or probability), and in some cases also underestimation of the mean failure probability of the system (e.g., if the components whose failure probabilities are dependent are functionally in parallel with each other); see Apostolakis and Kaplan (1981), Burmaster and Anderson (1994), and Kraan and Cooke (1997).

Historically, for reasons of computational tractability (among others), dependencies among random variables have often been either ignored, or else treated using unrealistic and simplistic assumptions such as perfect correlation. Fortunately, substantial progress is being made in modeling dependencies among components (and/or in the information about components). Two techniques, copulas (Nelsen, 1999) and Bayesian networks (Pearl, 1988), have become popular for specifying dependency relations. For example, copulas have been applied to model dependencies between opinions from different experts (Jouini and Clemen 1996; Lacke 1998) and between system failure rates during normal and accident conditions (Yi and Bier, 1998). They are also used extensively in financial risk analysis. Many explanations of the recent financial crisis emphasize the important role played by correlated financial risks—e.g., the idea that mortgage defaults, while perhaps approximately independent at ordinary times, may become highly correlated when housing prices are declining, and many houses in the same neighborhood may have resale values less than the remaining balances on their mortgages. Such correlations can be included in financial models using copulas (Frey et al., 2001). Bayesian networks use diagrams with arrows between variables indicating statistical dependencies among them and tables of conditional probability distributions for the value of each variable, given the values of the variables that point into it (if any) quantifying the statistical dependence among variables (Cox, 2015).

Example: analysis of accident precursors

To illustrate the use of copulas to analyze dependencies, we here consider how they can be used to help estimate the failure probabilities of critical safety systems in a nuclear power plant in the event of an accident. Fortunately, few nuclear power accidents have been observed, suggesting that analysts may wish to use more plentiful data on how often those safety systems have failed during routine testing. However, this data will clearly be only partially relevant to the probability of failure under accident conditions; for example, one might expect that many systems will have higher failure probabilities under accident conditions than during routine testing.

Yi and Bier (1998) show how copulas can be used to represent dependency between the system failure probabilities under normal versus accident conditions. This makes it possible to perform a Bayesian analysis applying data collected under normal conditions to help estimate system failure probabilities under accident conditions. Thus, for example, if routine testing showed a particular system to be much less reliable than was previously believed, this information could be used to update the expected failure probability of the system in the event of an accident.

Some alternatives to subjective prior distributions

Unlike classical statistical procedures, Bayesian analysis can be used in situations of sparse data, because subjective judgments and other non-statistical types of evidence can be used in Bayesian estimation, inference, and decision processes. However, with sparse data, the results of Bayesian analyses are often sensitive to the analyst's choice of prior probabilities for models and parameters (Small and Fischbeck, 1999). Hence, Bayesian methods can be more subjective and less readily accepted when data are sparse. Maximum-entropy distributions have sometimes been proposed to help solve this problem, as discussed in Bier et al. (1999), but have some significant disadvantages hindering their usefulness. In particular, maximum entropy can lead to significant problems even in some relatively simple examples. For example, if all we know about a random variable X is that it is bounded by 0 and 1, then a maximum-entropy distribution for it would be uniform between these limits. Of course, exactly the same reasoning presumably applies to X^2, but X and X^2 cannot both be uniformly distributed between 0 and 1.

Such limitations have raised interest in "robust" Bayesian methods (Rios Insua and Ruggeri, 2000) and other bounding approaches that update an entire class, family, or set of prior distributions with observed data, rather than just a single prior distribution. If the class is chosen carefully, the computational effort required to update all distributions in the class need not be substantially greater than for a single distribution. If all (or most) prior distributions in a suitably broad class give similar results, this can lead to greatly improved confidence in the results of the analysis.

In a similar spirit, probability bounds analysis (Ferson and Donald, 1998) propagates uncertainties (rather than choosing a single prior distribution for Bayesian updating). The analyst specifies bounds on the cumulative distribution functions of the various input parameters to a model, rather than selecting specific cumulative distributions. These bounds are then propagated through the model. The uncertainty propagation process, which again can be quite computationally efficient, yields valid bounds on the cumulative distribution function for the final result of the model (e.g., a risk level). This approach can take into account not only uncertainty about the probability distributions of the model inputs, but also uncertainty about their correlations and dependence structure. This is valuable, because correlations will often be more difficult to assess accurately than marginal distributions, and correlations of 1 or −1 among the input variables do not necessarily produce the most extreme possible distributions for the output variable(s) of interest (Ferson and Hajagos, 2006).

Characterizing model uncertainty

Copulas and maximum-entropy methods are mainly used to deal with uncertainties about the *parameters* and *input distributions* for particular models. However, *model uncertainties* (e.g., about which variables to include in a model, and/or the most appropriate *functional form* for a model) are frequently even more important in practice than input and parameter uncertainties. Failing to consider model uncertainties can lead to spuriously narrow statistical confidence intervals for parameter estimates and to spuriously high confidence in model-based predictions (Hoeting et al., 1999). Some researchers have therefore suggested assessing a probability distribution over multiple plausible models, by evaluating the consistency of the various models with the observed data (in much the same way as the likelihood function in Bayesian updating evaluates the consistency of various parameter values with observed data), and determining how much weight to put on each model based on its consistency with the data.

However, it is frequently not reasonable to attempt to estimate the probability that a given model is "correct," because, as Box and Draper (1987, p. 424) pointed out, "All models are wrong, some models are useful." For example, it seems highly implausible that any of the current models for estimating the probability of human error on a given task is close to being "correct" (because all are gross oversimplifications of the real world), nor can the current models be considered a collectively exhaustive set of possible models of human error. Moreover, some models may be intentionally conservative (e.g., for regulatory and/or screening purposes), or intentionally simplified (e.g., for computational tractability, or to yield qualitative insights). The fact that such models may be inconsistent with observed data does not necessarily invalidate their use for their intended purposes (Bier, 1995).

Finally, of course, more complex models, with larger numbers of parameters, will often fit the observed data well in many situations (subject to possible limitations of over-fitting), but may not always be preferable, if only for reasons of parsimony and/or generalizability. Thus, standard approaches for dealing with uncertainty probabilistically are often not well-suited for handling model uncertainty. Bayesian model averaging (BMA) was motivated largely by these challenges (Hoeting et al., 1999; Raftery and Zheng, 2003). BMA avoids basing all of one's conclusions on any single model if multiple models are more or less equally plausible. It also avoids giving high weight to models that are excessively complex if simpler ones give comparably good (or better) descriptions of the data, as measured by the likelihood of the data given a model. BMA generally performs reasonably well in practice (Hoeting et al., 1999; Raftery and Zheng, 2003).

An alternative that avoids assigning probabilities to individual models, "comprehensive uncertainty evaluation" (Brown, 1999), involves subjectively adjusting the probability distributions resulting from a particular model to try to take into account known weaknesses of the model (such as conservatisms, or risks that are not adequately modeled). This is consistent with subjective utility theory, and avoids some of the theoretical conundrums associated with assigning probabilities to models. Brown has applied this method (for example, to support regulatory decision-making for nuclear power plants), but it has not yet seen widespread application by other analysts in practice. Subjective assignments of probabilities to models can also be avoided in many situations where probability theory provides the required forms of distributions and/or useful bounds on the probable values of uncertain quantities. Table 1.1 summarizes some important classes of situations where probability theory prescribes distributions and bounds. Table 1.1 assumes familiarity with the various distributions mentioned, such as Poisson, Weibull, exponential, Gumbel, normal, and lognormal; see Ross (1995) for technical details of these distributions and topics. Googling the italicized topics in Table 1.1 will provide a host of web resources and authoritative references.

Note that many of these results can be applied even when the correct probability distributions are unknown or are only partly known, perhaps from statistical sampling or simulation modeling that provides estimates of means and variances. For example, the sums, maxima, and minima of repeated random samples from most distributions encountered in practice have asymptotic distributions (as the number of samples becomes large) that do not depend on the specific distribution being sampled. Thus, it is unnecessary to know the underlying "parent distribution" to quantify the distribution of these statistics, all of which are of interest in various risk analysis applications. Similarly, a variety of inequalities quantify how unlikely it is that a value sampled from a distribution will fall far from its expected value. Again, these bounds do not require detailed knowledge of the parent distribution. As a result, empirical data that give only limited information about a risky process may still be adequate to obtain useful quantitative bounds on risks of interest.

Example: the "rule of three" for negative evidence

Setting

People sometimes worry about events that *might* happen in theory, even though they *have not* (yet) happened in practice. How reassuring should one consider such "negative evidence" (i.e., absence of occurrences of a feared event, despite past opportunities for occurrence), bearing in mind the adage that "Absence of proof [of a hazard] is not proof of absence"? This can be an important topic when new technologies or poorly understood systems are involved, ranging from the Large Hadron Collider particle accelerator at CERN (which some feared might destroy the world by producing black holes), to the systems of interlocking safeguards that countries establish to try to protect against diseases such as bovine spongiform encephalitis ("mad cow" disease). We will illustrate how negative evidence (i.e., the observation that a feared event has not yet been observed) can be used to bound risk.

Problem

What if you have observed numerous test results, and never seen a failure (e.g., 100 successful tests of a valve with no failures, thousands of tests for a disease with no instances of disease found)? What can you say about the likelihood of failure or the prevalence of disease (since of course failures may be observed in future)?

Solution

A useful non-parametric confidence bound is based on the following "rule of three" (Chen and McGee, 2008): If an event that has probability p of occurring on each trial has not occurred in any of N independent trials (e.g., in a simple random sample of size N), then, with at least 95% confidence, its occurrence probability on each trial satisfies $p \leq 3/N$. Note that this bound does not require or assume any specific prior distribution for p, or any knowledge of the (possibly complex) processes that might give rise to the unobserved event.

Table 1.1 Selected asymptotic distributions and bounds from probability theory

Situation	*Key results and references*	*Examples*
Random occurrences	Can be modeled using a *Poisson distribution* (Ross, 1995). For example, the number of failures or other rare events observed in a fixed period of time is often approximately Poisson.	Bank robberies over time, car accidents on a stretch of highway, defects per square foot of material
Waiting times	Frequently modeled using an *exponential distribution* (Ross, 1995).	Time to failure of a component, time until arrival of a customer
Smallest value of a large number N of independent random variables; e.g., the time until the first of N components fails	• For large N, the smallest value has approximately a *Weibull distribution*, if all values are nonnegative. • If the N variables are unbounded (i.e., if they can take on arbitrary negative values, instead of being nonnegative), then the distribution of the smallest value approaches a *Gumbel distribution* as N gets large.	Failure time of a system with N components, if failure of any one component is sufficient to fail the system
Maximum value of a large number N of independent random variables	The largest value has a probability distribution that approaches a *Gumbel distribution* for large N (under certain reasonable conditions), or if those conditions are not satisfied, then a *generalized extreme value* distribution.	Maximum floods, rainfalls, traffic loads, insurance claims, bank deposits/ withdrawals
Sum of N independent random variables	The sum approaches a *normal distribution* for large N. *Central limit theorem*.	Total losses from N consecutive insurance claims; total time to use up N spare parts
Product of N independent random variables	The product approaches a *lognormal distribution* for large N.	Probability that all of many conditions hold; time to failure if further degradation occurs at a rate proportional to current degradation
X is a nonnegative random variable with mean $E(X)$	$\Pr[X \geq kE(X)] < 1/k$ for $k \geq 1$. *Markov's inequality* (http://mathworld.wolfram.com/ MarkovsInequality.html).	Random lifetime with unknown probability distribution, if mean is estimated from data
X is a random variable with mean $E(X)$ and variance σ^2	For any $k > 0$, $\Pr(\lvert X - E(X)\rvert \geq k\sigma) \leq 1/k^2$. In other words, values of the random variable X are unlikely to be many standard deviations away from the mean. *Chebyshev's inequality* (Ross, 1995; http://en.wikipedia.org/wiki/ Chebyshev's_inequality).	Probability of gain or loss exceeding a given limit (used in financial risk analysis)
Failure probability p, if no failures have been observed	If data consist of N trials with no successes, then there is approximately 95% confidence that p does not exceed $3/N$. *Rule of 3* (Chen and McGee, 2008; www.sinica.edu.tw/~jds/ JDS-401.pdf).	Upper bound for probability of a never-observed event

Risk characterization

The output of a PRA to support risk management decision-making is a characterization of the risk for each decision option being evaluated. Occasionally, the decision task is to identify an optimal risk management policy from a large set of possibilities, rather than to explicitly characterize the risks for each of a small number of alternatives. Then, simulation-optimization algorithms or other special-purpose techniques may be required. However, explicit comparison of risks from a few options is more usual, and is the main focus of this section.

"Risk" is usually defined in PRA as the frequency and severity of losses arising from operation of the designed system in its uncertain environment, including a specification of losses (i.e., which adverse consequences matter, and to whom). An effective display of risk shows how it is affected by different actions (e.g., different risk management decisions), and allows "drill-down" to view the risks to particular subpopulations, as well as the contributions of various different causes to the overall level of risk (see for example Table 1.2 below). For example, seeing how risks shift when risk-reducing measures are implemented would help managers identify the most effective measures. Uncertainty and sensitivity analysis are also essential to risk characterization, because they support estimates of the value of information.

Engineering vs. financial characterizations of "risk": why risk is not variance

The variance (or standard deviation) of the return on investment is widely used as a measure of risk in financial risk analysis, where mean-variance analysis is applied to calculate "efficient" frontiers and un-dominated portfolios, defined as those having maximum expected return for a given variance. Why, then, do health, safety, environmental, and reliability risk analysts insist on defining risk more flexibly, as being determined by probabilities *and* consequences, rather than simply by variances? A common critique in the theoretical decision analysis and financial economics literatures is that mean-variance analysis is valid only under restrictive and somewhat unrealistic conditions (e.g., if all risky prospects have normal distributions and utility functions are quadratic) (Baron, 1977; Markowitz, 1959). Mean-variance dominance and stochastic dominance relations for location-scale distributions do not coincide in general (Wong, 2006). Indeed, expected utility theory is inconsistent with models in which preferences

Table 1.2 Sample cause table for failure of reactor trip

Cause	*Frequency-failures per 10,000 demands*	*Comments*
• Common cause failures of reactor trip breakers	5.1	Occurred at Salem Nuclear Power Plant—*fortunately during testing!*
• Multiple independent failures of reactor trip breakers	0.39	
• Reactor trip system in test mode and one breaker fails	0.032	
Total	5.5	

Adapted from Pickard, Lowe and Garrick, Inc. (1983).

are determined by "moments" (mean, variance, etc.) of probability distributions (Brockett and Kahane, 1992). Variance is also inconsistent with proposed axioms for "coherent" or "rational" financial risk measures (Pedersen and Satchell, 1998). Moreover, empirical studies since the 1960s have demonstrated that real decision-makers pay attention to more than mean and variance in their choices among risky prospects (Jia et al., 1997).

A recent triumph of financial risk theory has been the definition and analysis of *coherent risk measures* (Artzner et al., 1999). These provide formulas for assigning numbers to risky prospects to ensure that comparisons of risk are logically consistent with each other. Older proposed measures of financial risk, including variance and value-at-risk, which reflects the probability of losing at least a specified amount, do not satisfy these axioms (Artzner et al., 1999; Pedersen and Satchell, 1998). Although coherent risk measures provide a substantial advance in methodologies for characterizing financial risks, they do not apply to risks that cannot be traded or diversified away via financial markets. Therefore, characterization of health, safety, environmental, and reliability risks in terms of probabilities or frequencies of different consequences, having different magnitudes or severities, is still the norm.

Communicating the results of PRAs

Risk communication (including both presenting the results of risk analyses to stakeholders, decision-makers, and other audiences, and listening to and actively eliciting their concerns so that they can be addressed in the risk analyses in the first place) facilitates the effective participation and interaction of technical experts, stakeholders, and decision-makers in risk management decisions and deliberations. There is an extensive body of literature on risk communication (including guidelines, survey results, experiments, and web resources on risk communication); see for example the literature reviews by Bier (2001a, 2001b), and the excellent guidance available from Hance et al. (1990).

Even more than PRA, risk communication is still an art rather than a science, but one that can be informed and improved by theory, experience, and experiments. Current challenges in risk communication include dealing with framing effects, communicating highly technical results to decision-makers who may not be intimately familiar with some of the methods used in the risk analysis, and building trust among affected stakeholders and members of the public more generally. Adding to the difficulty is the fact that communication and presentation styles that are most effective in accurately expressing the technical content of risk assessment findings may not always be those that invite and elicit public understanding, participation, and interaction.

Of course, technically accurate risk communication by itself is not sufficient to achieve other key goals of risk communication, such as changing people's behavior (Blaine and Powell, 2001), gaining their trust in the results of the analysis, or even giving them the information they need to make improved decisions. Rather, effective and persuasive communication about risks generally requires a concerted effort to build trust, gain and maintain credibility and legitimacy, and summarize relevant information simply and clearly (Bier, 2001b). Brevity, clarity, focus, candor, the use of cogent examples, and avoiding negative stereotypes may be crucial for communicating technical risks to non-specialist audiences in a way that ensures the message is heard and absorbed rather than tuned out or dismissed (e.g., Byrd and Cothern, 2000). Audience members generally respond not only (and sometimes not even primarily) to technical information about risks, but also to message framing, the source of the information, and the emotional style and assumed motives of the presenter in assessing the credibility of risk communication messages (Chartier and Gabler, 2001).

Using PRA results to improve risk management decision-making

Formal methods of decision analysis and optimization for uncertain systems have been extensively developed in operations research and systems engineering, and applied to both the design and the operation of complex engineering and industrial systems. Most share a simple common structure. The risk manager must choose from a set of feasible *controllable inputs* that influence a system's behavior. There are other facts and inputs (sometimes thought of as being selected by "nature" or "chance," and referred to as the state of the world) that cannot be directly selected by the risk manager, but still influence the system's behavior. The risk managers' acts and the state of the world together determine probabilities for different *consequences* (and for the system's next state). Finally, a *utility function* (Clemen and Reilly, 2005; Keeney and Raiffa, 1993) represents preferences for different consequences (or time streams of consequences) produced by the system. Various optimization algorithms and heuristics can be used to identify optimal (i.e., expected utility maximizing) or approximately optimal *acts*, given available information; or to identify optimal or approximately optimal *decision rules* (also called *policies*) that prescribe what acts to take based on the information available when decisions are made. Optimization algorithms for solving decision problems are constantly being refined and improved by ongoing research (Cox, 2015).

An important principle that cuts across many solution techniques for complex decision models is that *adaptive random sampling of potential solutions* is computationally tractable and finds "good" solutions to many problems that are too hard (e.g., too computationally complex) to solve using exact methods. Monte Carlo methods and related optimization "meta-heuristics" can estimate and optimize the expected utility of different acts or decision rules even for large and complex stochastic systems (Cox, 2015). Much as the mean value of a variable in a population can be estimated accurately from a random sample, regardless of the uncertainties and complexities of processes that created the population distribution of the variable, so the expected utility of a decision rule, policy, or act can often be estimated accurately using optimization algorithms that incorporate random sampling and adaptive improvement components.

Despite these advances in methods for decision analysis and optimization under uncertainty, in practice, formal decision analysis is seldom applied directly to make important risk management decisions. In part, this is because different participants may have different goals, different trade-offs among goals (e.g., minimizing average risk versus reducing inequities in the distribution of risks), and different tolerances for risk. In such cases, consensus utilities may not exist, and risk management decision-making requires not only analysis and deliberation (Stern and Fineberg, 1996), but also negotiation and compromise.

Even when decision analysis is not directly applied, however, its conceptual framework is still useful for organizing analysis and deliberation (Apostolakis and Pickett, 1998), separating beliefs from preferences, and identifying and resolving relevant conflicts and/or uncertainties about facts and values. Byrd and Cothern (2000) and Cox (2015) further discuss individual and group decision-making processes and frameworks for risk management decision-making.

Methods of risk management to avoid

Many decision processes focus on only a few aspects of the decision problem, such as minimizing the risk of the worst possible outcome, or maximizing the chance of a favorable outcome. However, such approaches have the disadvantage of not taking into account all aspects of a problem (such as whether the worst possible outcome is quite likely versus vanishingly rare, or how unfavorable the results are likely to be if a favorable outcome is not achieved). Decision theory (Clemen and Reilly, 2005; Keeney and Raiffa, 1993) recommends taking a more comprehensive view of the situation, weighting all outcomes by how likely they are. Just as one example, the popular idea of the precautionary principle violates this by focusing only on protecting the *status quo*; moreover, basing decisions on beliefs about what might constitute "precautionary" actions may not adequately take into account the likely consequences of alternative decision options. Likewise, it is important to consider the possibility that proposed risk-reduction actions might inadvertently *create* new risks (Bier, 1997; Dowell and Hendershot, 1997); ideally, risk characterization should provide risk managers with a balanced accounting of the adverse effects that a risk management intervention might *cause*, as well as those that it might *prevent*.

Models for security and infrastructure protection

Following September 11, 2001, there has been greatly increased interest in PRA for security, including protection of public and commercial buildings, water supply systems, and computer systems and software. Numerous researchers and practitioners have proposed the use of risk analysis in one form or another for homeland security (e.g., Garrick et al., 2004; Paté-Cornell and Guikema, 2002), especially for critical infrastructure (Apostolakis and Lemon, 2005; Ezell et al., 2001; Haimes et al., 1998). Recent work on reliability optimization (e.g., Levitin, 2003; Levitin et al., 2003; Levitin and Lisnianski, 2001) attempts to identify cost-effective risk-reduction strategies; for example, by optimizing physical separation of redundant components, or by allocating physical protection to various parts of a system (e.g., whether to harden the system as a whole, or individual components).

Much of the above work takes the "threat" against which systems are to be hardened to be static and unchanging. However, protection from intentional sabotage or terrorism differs from many other areas of risk management, because sabotage protection involves an intelligent adversary that can adapt in response to protective measures. Thus, reducing the vulnerability of some systems may cause adversaries to shift their attacks to other systems that have not yet been "hardened" to the same degree. Risk management in this context may therefore be better modeled as a "game" against an adversary—or, conversely, as a game between defenders, because security investment by one defender can have either positive or negative effects on the threats faced by other defenders (Kunreuther and Heal, 2003). Game theory was once viewed as being of little relevance for practical risk management decision-making. However, several recent developments have started to change that view. These include not only increased interest in terrorism, homeland security, and critical infrastructure protection (which can be effectively modeled using game theory), but also increased interest in risk-informed regulation (which can be viewed as a game between a regulator and a regulated firm; Bier and Lin, 2013). As a result of such developments, game theory is becoming an important tool in a variety of application areas related to risk, as discussed below.

A large body of work now applies game theory to security (Bier and Azaiez, 2009), much of it by economists (e.g., Arce and Sandler, 2001; Enders and Sandler, 2004; Frey and Luechinger, 2003; Keohane and Zeckhauser, 2003; Lakdawalla and Zanjani, 2005). Much of this work is intended to inform policy-level decisions; e.g., by clarifying the relative merits of public versus private funding of defensive investments, or deterrence versus other protective measures. Recently, efforts have begun to focus more on operational risk management decisions, such as deciding how much defensive investment to allocate to particular assets (e.g., O'Hanlon et al., 2002), and have more of a risk analysis flavor (e.g., taking the success probabilities of potential attacks into account); see for example Bier et al. (2005) and Woo (2002). As one illustration, Hausken (2002) has applied game theory to study allocation of resources to ensuring component (and hence system) reliability in situations where different individuals or groups are responsible for the reliability of different components; in this situation, system reliability is viewed as a "public good," where agents responsible for the reliability of one component might attempt to "free ride" on investments in the reliability of other components in that system (e.g., postponing needed reliability enhancements in the hopes that some other agent will implement such improvements instead).

Conclusions

This chapter has surveyed methods and concepts for PRA and decision-making in engineered systems. Although modeling of uncertain systems has been tremendously enabled by recent advances, PRA still poses many challenges. Technical challenges remain in how best to construct useful (and at least approximately valid) models of systems and their environments from engineering knowledge and data, and in identifying optimal or near-optimal risk management policies. Communicating the results effectively and using them to guide improved decision-making by multiple parties (e.g., teams of stakeholders) also poses practical questions that go beyond the framework of single-person decision theory. If the environment in which a system operates includes intelligent adversaries, then insights from novel methods (e.g., game-theoretic principles) may be needed to ensure that risk-reduction strategies are effective and cost-effective. These challenges are likely to stimulate further advances in both the theory and practice of decision sciences for engineering risk analysis. However, even in light of those challenges, PRA provides a clear method of identifying and analyzing the risks of extreme events, and a flexible tool for managing safety (Bley et al., 1992).

References

Apostolakis, G. and Kaplan, S. (1981). "Pitfalls in risk calculations." *Reliability Engineering and System Safety*, 2, 135–145.

Apostolakis, G. E. and Lemon, D. M. (2005). "A screening methodology for the identification and ranking of infrastructure vulnerabilities due to terrorism." *Risk Analysis*, 25, 361–376.

Apostolakis, G. E. and Pickett, S. E. (1998). "Deliberation: integrating analytical results into environmental decisions involving multiple stakeholders." *Risk Analysis*, 18, 621–634.

Arce, M., Daniel, G., and Sandler, T. (2001). "Transnational public goods: Strategies and institutions." *European Journal of Political Economy*, 17, 493–516.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). "Coherent measures of risk." *Mathematical Finance*, 9, 203–228.

Barlow, R. E. (1998). *Engineering Reliability*. American Statistical Association and Society for Industrial and Applied Mathematics (ASA-SIAM), ASA-SIAM Series on Statistics and Applied Probability. Philadelphia, PA: ASA-SIAM.

Barlow, R. E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing: Probability Models*. New York: Holt, Rinehart and Winston.

Baron, D. P. (1997). "On the utility-theoretic foundation of mean-variance analysis." *The Journal of Finance*, 32(5), 1683–1697.

Bier, V. M. (1995). "Some illustrative examples of model uncertainty." In *Model Uncertainty: Its Characterization and Quantification* edited by A. Mosleh, N. Siu, C. Smidts, and C. Lui (pp. 93–100). College Park, MD: University Press of Maryland.

Bier, V. M. (1997). *Illusions of safety*. Paper presented at the Workshop on Organizational Analysis in High Hazard Production Systems: An Academy/Industry Dialogue, Dedham, MA. http://www.chpra.wisc.edu/pdfs/IllusionsofSafety.pdf.

Bier, V. M. (2001a). "On the state of the art: Risk communication to decision-makers." *Reliability Engineering and System Safety*, 71, 151–157.

Bier, V. M. (2001b). "On the state of the art: Risk communication to the public." *Reliability Engineering and System Safety*, 71, 139–150.

Bier, V. M. and Azaiez M. N. (Eds.) (2009). *Game Theoretic Risk Analysis of Security Threats*. New York: Springer.

Bier, V. M., Cox and L. A., Jr. (2007). "Probabilistic risk analysis for engineered systems." In *Advances in Decision Analysis*, edited by W. Edwards, R. Miles, and D. von Winterfeldt (pp. 279–301). Cambridge, UK: Cambridge University Press. http://www.cambridge.org/us/catalogue/catalogue.asp?isbn=0521682304

Bier, V. M., Haimes, Y. Y., Lambert, J. H., Matalas, N. C. and Zimmerman, R. (1999). "Assessing and managing the risk of extremes." *Risk Analysis*, 19, 83–94.

Bier, V. M. and Lin, S.-W. (2013). "Should the model for risk-informed regulation be game theory rather than decision theory?" *Risk Analysis*, 33, 281–291.

Bier, V. M., Nagaraj, A., and Abhichandani, V. (2005). "Optimal allocation of resources for defense of simple series and parallel systems from determined adversaries." *Reliability Engineering and System Safety*, 87, 313–323.

Blaine, K., and Powell, D. (2001). "Communication of food-related risks." *AgBioForum*, 4, 179–185.

Bley, D., Kaplan, S., and Johnson, D. (1992). "The strengths and limitations of PSA: Where we stand," *Reliability Engineering and System Safety*, 38, 3–26.

Bobbio, A., Portinale, L., Minichino, M., and Ciancamerla, E. (2001). "Improving the analysis of dependable systems by mapping fault trees into Bayesian Networks." *Reliability Engineering and System Safety*, 71, 249–260.

Borgonovo, E., Smith, C. L., Apostolakis, G. E., Deriot, S. and Dewailly, J. (2000). "Insights from using influence diagrams to analyze precursor events." In *Proceedings of PSAM 5, Probabilistic Safety Assessment and Management*, edited by S. Kondo and K. Furuta (pp. 1801–1807). Tokyo: Universal Academy Press.

Box, G. E. P. and Draper, N. R. (1987). *Empirical Model-Building and Response Surfaces*. New York: John Wiley & Sons.

Brockett, P. L., Kahane, Y. (1992). "Risk, Return, Skewness and Preference." *Management Science*, 38(6), 851–866.

Brown, R. (1999). *Using Soft Data to Make "Probabilistic Risk Assessments" Realistic*. Retrieved July 29, 2005, http://fisher.osu.edu/~butler_267/DAPapers/WP000002.pdf

Bucklew J. A. (2004). *Introduction to Rare Event Simulation*. New York: Springer.

Burmaster, D. and Anderson, P. D. (1994). "Principles of good practice for the use of Monte Carlo techniques in human health and ecological risk assessments." *Risk Analysis*, 14, 477–481.

Byrd, D. M., and Cothern, C. R. (2000). *Introduction to Risk Analysis: A Systematic Approach to Science-Based Decision Making*. Houston: ABS Group.

Chartier, J., and Gabler, S. (2001). *Risk Communication and Government: Theory and Application for the Canadian Food Inspection Agency*. Ottawa: Canadian Food Inspection Agency.

Chen, Z. and McGee, M. (2008). "A Bayesian approach to zero-numerator problems using hierarchical models." *Journal of Data Science*, 6, 261–268.

Clemen, R. and Reilly, T. (2005). *Making Hard Decisions: Introduction to Decision Analysis*. Pacific Grove, CA: Duxbury Press.

Cox, L. A., Jr. (Ed.) (2015). *Breakthroughs in Decision and Risk Analysis*. Hoboken, NJ: Wiley.

Defriend, S., Dejmek, M., Porter, L., Deshotels, B., and Natvig, B. (2008). "A risk-based approach to flammable gas detector spacing." *Journal of Hazardous Materials*, 159, 142–151.

Dowell, A. M. and Hendershot, D. C. (1997). "No good deed goes unpunished: Case studies of incidents and potential incidents caused by protective systems." *Process Safety Progress*, *16*, 132–139.

Enders, W., and Sandler, T. (2004). "What do we know about the substitution effect in transnational terrorism?" In *Research on Terrorism: Trends, Achievements, Failures*, edited by A. Silke (pp. 119–137). London: Frank Cass.

Ezell, B. C., Haimes, Y. Y., and Lambert, J. H. (2001). "Cyber attack to water utility supervisory control and data acquisition (SCADA) systems." *Military Operations Research*, 6, 23–33.

Ferson, S. and Donald, S. (1998). "Probability bounds analysis." In *Probabilistic Safety Assessment and Management*, edited by A. Mosleh and R. A. Bari (pp. 1203–1208). New York: Springer-Verlag,

Ferson, S. and Hajagos, J. G. (2006). "Varying correlation coefficients can underestimate uncertainty in probabilistic models." *Reliability Engineering and System Safety*, 91, 1461–1467.

Frey, B. S. and Luechinger, S. (2003). How to fight terrorism: Alternatives to deterrence. *Defence and Peace Economics*, 14, 237–249.

Frey, R., McNeil, A.J., and Nyfeler, M. A. (2001). "Copulas and credit models." *Risk* 10, 111–114.

Garrick, B. J., Hall, J. E., Kilger, M., McDonald, J. C. et al. (2004). "Confronting the risks of terrorism: Making the right decisions." *Reliability Engineering and System Safety*, 86, 129–176.

Haimes, Y. Y., Matalas, N. C., Lambert, J. H., Jackson, B. A., and Fellows J. F. R. (1998). "Reducing vulnerability of water supply systems to attack." *Journal of Infrastructure* Systems, 4, 164–177.

Hance, B. J., Chess, C. and Sandman, P. M. (1990). *Industry Risk Communication Manual*. Boca Raton, FL: CRC Press/Lewis Publishers,

Hausken, K. (2002). "Probabilistic risk analysis and game theory." *Risk Analysis*, 22, 17–27.

Hoeting, J.A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). "Bayesian model averaging: A tutorial." *Statistical Science*, 14(4), 382–401. http://mpdc.mae.cornell.edu/Courses/UQ/2676803.pdf

Howard, R. A. (1971). Dynamic Probabilistic Systems: Vol. 1: Markov Models; Vol. II: Semi-Markov and Decision Processes. New York: Wiley.

Issel, L.M. and Narasimha, K. M. (2007). "Creating complex health improvement programs as mindful organizations: From theory to action." *Journal of Health Organization and Management*, 21(2/3), 166–183.

Jia J., Dyer, J. S., and Butler, J. C. (1999). "Measures of perceived risk." *Management Science*, 45(4), 519–532. http://fisher.osu.edu/~butler_267/DAPapers/WP970005.pdf

Kahneman, D., Ritov, I.and Schkade, D. (1999). "Economic preferences or attitude expressions? An analysis of dollar responses to public issues." *Journal of Risk and Uncertainty*, 19, 203–235.

Kaplan, S. and Garrick, B. J. (1981). "On the quantitative definition of risk." *Risk Analysis*, 1, 11–27.

Keeney, R. L. and Raiffa, H. (1993). *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*. New York: Cambridge University Press.

Keohane, N. O. and Zeckhauser, R. J. (2003). "The ecology of terror defense." *Journal of Risk and Uncertainty*, 26, 201–229.

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V. and Fotiadis, D. I. (2014). "Machine learning applications in cancer prognosis and prediction." *Computational and Structural Biotechnology Journal*, 13, 8–17.

Kraan, B. and Cooke, R. (1997). "The effect of correlations in uncertainty analysis: Two cases." In *Technical Committee Uncertainty Modeling: Report on the Benchmark Workshop Uncertainty/ Sensitivity Analysis Codes*, edited by R. Cooke. Delft, the Netherlands: European Safety and Reliability Association.

Kunreuther, H. and Heal, G. (2003). "Interdependent security." *Journal of Risk and Uncertainty* 26, 231–249.

Lacke, C. (1998). *Decision Analytic Modeling of Colorectal Cancer Screening Policies*. Unpublished doctoral dissertation. North Carolina State University.

Lakdawalla, D., and Zanjani, G. (2005). Insurance, self-protection, and the economics of terrorism. *Journal of Public Economics* 2005; 89: 1891–1905.

Lave, L. B. and Balvanyos, T. (1998). "Risk Analysis and Management of Dam Safety." *Risk Analysis*, 18, 455–462.

Lee, P. M. (2004). *Bayesian Statistics: An Introduction*, 3rd edn. London: Arnold.

Jouini, M, and Clemen, R. T. (1996). "Copula models for aggregating expert opinions." *Operations Research*, 44, 444–457.

Leung, M., Lambert, J. H. and Mosenthal, A. (2004). "A risk-based approach to setting priorities in protecting bridges against terrorist attacks." *Risk Analysis*, 24, 963–984.

Leveson, N. G. (2004). "A systems-theoretic approach to safety in software-intensive systems." *IEEE Transactions on Dependable and Secure* Computing, 1(1), 66–86.

Levitin, G. (2003). "Optimal multilevel protection in series-parallel systems." *Reliability Engineering and System Safety*, 81, 93–102.

Levitin, G., Dai, Y., Xie, M., and Poh, K. L. (2003). "Optimizing survivability of multi-state systems with multi-level protection by multi-processor genetic algorithm." *Reliability Engineering and System Safety*, 82, 93–104.

Levitin, G. and Lisnianski, A. (2001). "Optimal separation of elements in vulnerable multi-state systems." *Reliability Engineering and System Safety*, 73, 55–66.

Lv, Y., Huang, G. H., Guo, L., Li, Y. P., Dai, C., Wang, X. W. and Sun, W. (2013). "A scenario-based modeling approach for emergency evacuation management and risk analysis under multiple uncertainties." *Journal of Hazardous Materials*, 246–247, 234–244.

Markowitz, H. M. (1959). *Portfolio Selection: Efficient Diversification of Investments*. New York: John Wiley & Sons.

McGill, W. L., Ayyub, B. M. and Kaminskiy, M. (2007). "Risk analysis for critical asset protection," *Risk Analysis*, 27, 1265–1281.

Mili, L., Qiu, Q. and Phadke, A. G. (2004). "Risk assessment of catastrophic failures in electric power systems," *International Journal of Critical Infrastructures*, 1, 38–63.

Montazeri, M., Montazeri, M., Montazeri, M. and Beigzadeh, A. (2016). "machine learning models in breast cancer survival prediction," *Technology and Health Care*, 24, 31–42.

Montibeller, G. and Franco, A. (2007). "Decision and risk analysis for the evaluation of strategic options." In *Supporting Strategy: Frameworks, Methods and Models* edited by F. A. O'Brien and R. G. Dyson (pp. 251–284). Chichester, UK: Wiley.

Mosleh, A., Rasmuson, D. and Marshall, F. M. (1998). *Guidelines on Modeling Common Cause Failures in Probabilistic Risk Assessments*. Washington, DC: U.S. Nuclear Regulatory Commission,

Nelsen, R. B. (1999). *An Introduction to Copulas*. New York: Springer-Verlag.

O'Hanlon, M., Orszag, P., Daalder, I., Destler, M., Gunter, D., Litan, R., and Steinberg J. (2002). *Protecting the American Homeland*. Washington, DC: Brookings Institution.

Ólafsson, S. and Kim, J. (2002). "Simulation optimization." In *Proceedings of the 2002 Winter Simulation Conference,* edited by E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes (pp. 79–84). Sandiago, CA: IEEE.

Paté-Cornell, E. and Cox, L. A., Jr. (2014). "Improving risk management: From lame excuses to principled practice." *Risk Analysis*, 34, 1228–1239.

Paté-Cornell, M. E. and Fischbeck, P. S. (1993). "Probabilistic risk analysis and risk-based priority scale for the tiles of the space shuttle." *Reliability Engineering and System Safety*, 40, 221–238.

Paté-Cornell, E., and Guikema, S. (2002). "Probabilistic modeling of terrorist threats: A systems analysis approach to setting priorities among countermeasures." *Military Operations Research*, 7, 5–20.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems*. San Francisco, CA: Morgan Kaufmann.

Perrow, C. (1984). *Normal Accidents: Living with High-Risk Technologies*. New York: Basic Books.

Pedersen, C. S. and Satchell, S. E. (1999). "Choosing the right risk measure: A survey." http://citeseer.ist.psu.edu/520923.html

Phimister, J. R., Bier, V. M. And Kunreuther, H. C. (Eds.) (2004). *Accident Precursor Analysis and Management: Reducing Technological Risk through Diligence*. Washington, DC: National Academies Press.

Pickard, Lowe and Garrick, Inc. (1983). *Seabrook Station Probabilistic Safety Assessment*. Prepared for Public Service Company of New Hampshire and Yankee Atomic Electric Company.

Plett, G. L. (2003). "Adaptive inverse control of linear and nonlinear systems using dynamic neural networks," *IEEE Transactions on Neural Networks*, 14, 360–376.

Raftery, A. E. and Zheng, Y. (2003). "Discussion: Performance of Bayesian model averaging." *Journal of the American Statistical Association*, 98, 931–938.

Rios Insua, D. and Ruggeri, F. (Eds.) (2000). *Robust Bayesian Analysis*. New York: Springer-Verlag.

Ross, S. M. (1995). *Stochastic Processes*, 2nd edn. New York: Wiley.

Rubinstein, R. Y.and Kroese, D. P. (2004). *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning*. New York: Springer-Verlag.

Small, M. J. and P. S. Fischbeck (1999). "False precision in Bayesian updating with incomplete models." *Journal of Human and Ecological Risk Assessment*, 5(2), 291–304.

Smith, D. J. (2005). *Reliability, Maintainability and Risk: Practical Methods for Engineers Including Reliability Centered Maintenance and Safety-Related Systems*, 7th edn. New York: Elsevier.

Stern, P. C., and Fineberg, H. V. (Eds.) (1996). *Understanding risk: Informing decisions in a democratic society*. Washington, DC: National Academy Press.

Sunstein, C. R. (2002). "Probability neglect: Emotions, worst cases, and law." *Yale Law Journal*, 112, 61–107.

Till, J. E., Rood, A. S., Garzon, C. D., and Lagdon, R., Jr. (2014). "Comparison of the MACCS2 atmospheric transport model with Lagrangian Puff models as applied to deterministic and probabilistic safety analysis." *Health Physics*, 107, 213–230.

U.S. Nuclear Regulatory Commission (1975). Reactor Safety Study—An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants, WASH-1400, NUREG-75/014.

Weick, K.E. and Sutcliffe, K. M. (2007). *Managing the Unexpected: Resilient Performance in an Age of Uncertainty*, 2nd edn. John Wiley & Sons.

Woo, G. (2002). "Quantitative terrorism risk assessment." *Journal of Risk Finance*, 4(1), 7–14. http://www.rms.com/newspress/quantitative_terrorism_risk_assessment.pdf

Wong, W-K. (2006). Stochastic dominance theory for location-scale family. *Journal of Applied Mathematics and Decision Sciences*, 2006, Article ID 82049, Pages 1–19. http://dx.doi.org/10.1155/JAMDS/2006/82049

Wurbs, R. A. (2005). *Comparative Evaluation of Generalized River/Reservoir System Models*. College Station, TX: Texas Water Resources Institute.

Yi, W., and Bier, V. M. (1998). "An application of copulas to accident precursor analysis," *Management Science*, 44, S257–S270.

2 The meaning of black swans

Terje Aven

Introduction

The black swan metaphor is intuitively appealing and has been very popular for illustrating the idea of surprising events and outcomes in a risk context (Aven 2013b). The black swan term was first introduced by the Latin poet Juvenal, who wrote "rara avis in terris nigroque simillima cygno" (a rare bird upon earth, and exceedingly like a black swan). However, according to Hammond (2009) this was imaginative irony. Juvenal's phrase was a common way of speaking in sixteenth-century London, as an expression of something impossible. Up to that point in time, all observed swans in the Old World had been white. Taleb (2007, p. xvii), writes: "Before the discovery of Australia people in the Old World were convinced that all swans were white, an unassailable belief as it seemed completely confirmed by empirical evidence." Then in 1697 a Dutch expedition to Western Australia, led by Willem de Vlarningh, discovered black swans on the Swan River, and the black swan concept was developed to suggest not only something extremely rare, but also that a perceived impossibility might later be disproven: a logical fallacy, in the sense that, if one does not know about something, it is therefore impossible. Taleb (2007) comments that, in the nineteenth century, John Stuart Mill used the black swan logical fallacy as a new term to identify falsification. Mill wrote: "No amount of observations of white swans can allow the inference that all swans are white, but the observation of a single black swan is sufficient to refute that conclusion." It became a classic case in elementary philosophy (Hammond 2009).

In 2007, Nassim Taleb further defined and popularised the black swan concept in his book, *The Black Swan* (Taleb 2007). Taleb provides a number of examples of black swans:

> Just imagine how little your understanding of the world on the eve of the events of 1914 would have helped you guess what was to happen next. (Don't cheat by using the explanations drilled into your cranium by your dull high school teacher.) How about the rise of Hitler and the subsequent war? How about the precipitous demise of the Soviet bloc? How about the rise of Islamic fundamentalism? How about the spread of the Internet? How about the market crash of 1987 (and the more unexpected recovery)? Fads, epidemics, fashion, ideas, and the emergence of art genres and schools. All follow these Black Swan dynamics. Literally, just about everything of significance around you might qualify.
>
> (Taleb 2007, p. x)

Taleb makes an interesting comment on the September 11 event:

> Think of the terrorist attack of September 11, 2001: had the risk been reasonably *conceivable* on September 10, it would not have happened. If such a possibility were deemed worthy of attention, fighter planes would have circled the sky above the twin towers, airplanes would have had locked bulletproof doors, and the attack would not have taken place, period. Something else might have taken place. What? I don't know.
>
> (Taleb 2007, p. x)

These comments raise a number of questions. What actually is a black swan in a professional risk context? What does it mean that risk is reasonably conceivable? Can risk assessment identify black swans? And how shall we confront such events, i.e. how shall we manage the risk related to black swans? These are issues discussed in this chapter. We review and discuss recent work carried out to explore the concept of black swans and on how to understand, assess, and manage this type of risk.

It is obvious that the traditional probability-based approach to risk assessment and management does not provide the answer to meeting black swan types of events. Instead, "robust/resilient approaches," are often referred to as methods that are effective in meeting hazards/threats, surprises, and the unforeseen (Hollnagel et al. 2006; Weick and Sutcliffe 2007; Aven 2015). These approaches cover cautionary measures such as designing for flexibility (meaning that it is possible to utilise a new situation and adapt to changes); implementing barriers; improving the performance of barriers by using redundancy, testing, etc.; and applying quality control/assurance. The term "resilient" covers ways to enhance the ability of organisations to recognise, adapt to, and absorb variations, changes, disturbances, disruptions, and surprises (Hollnagel et al. 2006). We may also include the concept of antifragility (Taleb 2012): the key being to love randomness, variation, and uncertainty to some degree, and thus also errors. Just as our bodies and minds need stressors to be in top shape and improve, so do other activities and systems. In practice, for example in relation to industrial activities such as nuclear and oil and gas, the standard risk management approach represents a balance between probability-based risk management and robust/resilient methods.

It is a common perception that the challenge in dealing with surprises and black swans lies in the fact that they are beyond the sphere of probability and risk (e.g. Gross 2010). However, the present chapter is based on the thesis that the issue of confronting such events is at the core of risk management, when risk is properly understood. A suitable conceptualisation for this has been established (e.g. Aven and Renn 2010; Aven et al. 2014); see also the third section of this chapter; now the challenge is to develop adequate types of risk assessments and management policies. Before we look into different types of black swans and how to confront this type of risk, let us briefly look at some criticism that has been levelled against the use of this concept.

Critique against the use of the black swan concept

Some scholars are sceptical about the use of the black swan concept, as discussed for example by Aven (2013b). Professor Dennis Lindley, one of the strongest advocators of the Bayesian approach to probability, statistics, and decision making, made his view very clear in a review of Taleb's black swan book (Lindley 2008): Taleb talks nonsense. Lindley lampoons Taleb's distinction between the lands of Mediocristan and Extremistan, the former capturing the placid randomness as in the tossing of a coin, and the latter covering the randomness that provides the unexpected and extreme outcomes (the black swans). Lindley provides an example of a sequence of independent trials with a constant unknown probability (chance) of success—clearly an example of Mediocristan. Each trial is to be understood as a swan, and success as a white swan. Using simple probability calculus rules, Lindley shows that a black swan is almost certain to arise if one is to see a lot of swans, although the probability that the next swan observed is white, is nearly one. Lindley cannot be misunderstood: "Sorry, Taleb, but the calculus of probability is adequate for all cases of uncertainty and randomness" (Lindley 2008, p. 42).

Let us look more closely at the details. We consider a sequence of independent trials with a constant unknown success chance. Lindley assumes a prior probability distribution for this chance, a uniform distribution over the interval [0,1]. This means that Lindley tacitly has assumed that there is a zero probability that all swans are white—there is a proportion of swans out there that are non-white, i.e. black. Of course, then the probabilistic analysis will show that, when considering a sufficient number of swans, some black ones will be revealed; see Aven (2013b) for the formal proof. Through the assumptions made, the analyst has removed the main uncertainty aspect. In real life, we cannot exclude the possibility that all swans are white. The uncertainty about all swans being white is a central issue here, and Lindley has concealed it in the assumptions. This is the challenge raised by many authors: the probability-based approach to treating the risk and uncertainties is based on assumptions that could conceal critical assumptions and therefore provides a misleading description of the possible occurrence of future events. Let us reconsider Lindley's example to allow for a positive probability that all swans are white.

Let us assume that there are only two possibilities: the proportion p of white swans is either 1 or 0.99. Suppose the analysts assign prior probabilities to these values as 0.2 and 0.8, respectively. Now suppose that the analysts have observed n swans and they are all white, what then is their posterior probability for the next m swans to all be white (n and m being large numbers)? Using Bayes' formula in the standard way, it can be shown that this probability is close to one, i.e. the probability of a black swan occurring is very small, in contrast to what Lindley found in his analysis. See Aven (2013b) for the details.

This example shows the importance of the assumptions made for the probabilistic study. Depending on these assumptions, we arrive at completely different conclusions about the probability of a black swan occurring.

Lindley's calculations also fail to reflect the essence of the black swan issue in another way. In real life, the definition of a probability model and chances cannot always be justified, as it presumes some stability of situations and experiments similar to the one analysed (Aven 2013b). Lindley's set-up is the common framework used in both traditional statistics and Bayesian analysis. Statisticians and others often simply presume the existence of this type of framework, and the elements of surprise that Taleb and others are concerned about fall outside the scope of the analysis. This is the key problem of the probability-based approach to risk assessment, and a possible interpretation of Taleb's work is a critique of the lack of will and interest among statisticians and others to see beyond this framework when studying risk.

In relation to this discussion, one may question why it is important to clarify the meaning of a black swan. Why should we try to agree on the definition of a black swan as a scientifically based term in this field?

As a response to such questions, and following the reflections made in Aven (2013b), it should be highlighted that what is important is not really the term but the related analysis and management—the structure developed to understand and analyse the related features of risk and uncertainties in a risk assessment and risk management context. Nonetheless, the concept of the black swan exists—the idea has gained a lot of attention and is a hot topic in many forums that discuss risk and safety/security. As a scientific and professional community, we cannot just ignore this. We need to provide perspectives and guidance on what this concept is expressing. We need to place this concept in the frameworks that the risk and safety/security fields have developed over the years. Furthermore, the risk and safety/security fields need suitable concepts for reflecting this type of phenomena. The popularity of the black swan metaphor clearly demonstrates this, but, it should be added, also from a strictly professional point of view, there is a need for concepts that reflect what a "surprise" really means in this context. These concepts cannot and should not be limited by the probability-based perspective, as the phenomena that we are trying to characterise extend beyond this paradigm. In the extended non-probability setting we need to develop proper terminology, and the black swan concept can be seen as a useful contribution to this end.

Using the black swan concept, the present author has noticed increased interest and enthusiasm for discussing risk issues. In addition, studying the black swan concept will provide new insights into the risk and safety/security fields regarding the links between risk, probability, and uncertainties. The present author's research on this topic is based on the conviction that there is a need for strengthening the foundation of these fields by providing new insights into the relationship between surprising events, risk, probability, and uncertainty. For these fields, as for all other scientific fields (disciplines), it is essential that the conceptual basis is solid.

The concept of a black swan. Different types of black swans

Taleb (2007) refers to a black swan as an event with the following three attributes: firstly, it is an outlier, as it lies outside the realm of regular expectations, because nothing in the past can convincingly point to its possibility. Secondly, it carries an extreme impact. Thirdly, in spite of its outlier status, human nature makes us concoct explanations for its occurrence after the fact, making it explainable and predictable. Other definitions of black swans have also been suggested. Aven (2013b) refers to a black swan as a surprising extreme event relative to one's knowledge. Here, knowledge is interpreted as justified beliefs. It is typically based on data and information. A risk analysis group may have strong knowledge about how a system works and can provide strong arguments as to why it will not fail over the next year, but it cannot know for sure whether or not it will in fact fail. Nobody can. A group of experts can believe that a system will not be able to withstand a specific load. Their belief is based on data and information, modelling and analysis, but they can be wrong. The system could be able to withstand this load. As a third example, is it true that the (frequentist) probability of a fatal accident in a process plant is higher than 1×10^{-4}, even if a large amount of data shows this and all experts agree on it? Leaving aside the issue that such probabilities may be difficult to define, we cannot say that we have a true belief that this probability is so large, as the probability is unknown. We can have a strong belief, and we can even introduce probabilities (subjective) to express the strength of the belief.

Following this line of thinking, knowledge cannot of course be objective, as a belief is someone's belief. In general, knowledge then needs to be considered as subjective or at best inter-subjective among people, for example experts.

From such a view, the term "justified" becomes critical. Philosophers and others have discussed the issue since ancient times. We link it to being a result of a reliable process, a process that generally produces true beliefs. It applies to the justification of a specific statement by an individual and broad justifications of scientific theses (Aven 2014), for example linked to the assignment of a subjective (knowledge-based, judgemental) probability, quantified risk assessments, and the science of risk analysis. The degree of reliability will always be an issue for debate.

In Aven and Krohn (2014) and Aven (2015), three main types of black swan events are identified based on this definition (see Figure 2.1):

a) events that were completely unknown to the scientific environment (unknown unknowns);
b) events not on the list of known events from the perspective of those who carried out a risk analysis (or another stakeholder), but known to others (unknown knowns—unknown events to some, known to others);
c) events on the list of known events in the risk analysis but judged to have negligible probability of occurrence, and thus not believed to occur.

The term "black swan" is used to express any of these types of events, tacitly assuming that it carries an extreme impact.

The first category of black swan type of events (a) is the extreme—the type of event is unknown to the scientific community. A good example is the effects of the thalidomide drug (Smithells and Newman 1992). The drug was introduced in 1957, and not long afterwards children were observed with gross limb malformations of an unusual form.

The second type of black swans (b) is events that are not captured by the relevant risk assessments, either because we do not know them or we have not made a sufficiently thorough consideration. If the event then occurs, it was not foreseen. If a more thorough

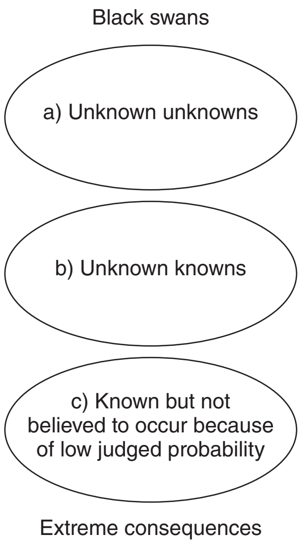


Figure 2.1 Three different types of black swans in a risk context (based on Aven and Krohn 2014).

risk analysis had been conducted, the event could have been identified. The September 11 attack is a good example of this type of black swan.

The third category of black swans comprises events that occur despite the fact that the occurrence probability was judged to be negligible. The events are known, but considered so unlikely that they are ignored—they are not believed to occur, and cautionary measures are not implemented. An example is the event that an underwater volcano eruption occurs in the Atlantic Ocean resulting in a tsunami affecting, for example, Norway. The events are on the list of risk sources and hazards but then removed as their probability is judged as negligible. Their occurrence will come as a surprise. The tsunami that destroyed the Fukushima Daiichi nuclear plant was similarly removed from the relevant risk lists due to the judgment of negligible probability.

The black swans' "surprising" aspect must always be understood in relation to by whom and when. We consider an activity, for instance the operation of an offshore installation, at a given future time period, for example the next 10 years. Let C denote the consequences of the activity in relation to the values we are concerned about (life, health, environment, assets). What C will be is unknown to us now at the time *s*; there is risk present. We assume that a risk assessment of the activity has been conducted at the time *s*. Time goes by, and C is realised, hopefully without a major accident occurring. But let us assume that such an accident actually occurs. Think about the Macondo accident in 2010 as an example. The accident is a result of a combination of conditions and events occurring and comes as a surprise to those involved in the management of the activity. In the Macondo example, this combination includes (National Commission 2011):

- erroneous assessments of the results of pressure tests;
- failure to identify that the formation fluid penetrated the well, in spite of the fact that log data showed that this was the case;
- the diverter system was unable to divert gas;
- the cutting valve (Blind Shear Ram, BSR) in the blowout preventer (BOP) did not seal the well.

The accident is a black swan for them. It came as a surprise that such a sequence of events could occur; they had not foreseen such a scenario. It is a black swan of type (b). If they had thought of it, they would have concluded that it was so unlikely that one could ignore it. It would have been a black swan of type (c).

Now let us take a macro perspective—looking at a considerable number of such activities, for example the whole oil industry. Risk is now associated with the occurrence of any major accident in the industry; where and how the accident event occurs is not the issue. Again, a risk assessment is performed. It is concluded that there is quite a high probability that such an accident could occur. Consequently, one cannot say that it is a black swan if such an event does in fact occur. From a macro perspective, a realistic study would state that we must expect that a major accident will occur somewhere in the next ten years. However, there is no law that says that it will actually happen. We are not subject to destiny or fate. Each unit (installation, company, or organisation) works hard to prevent such an accident from occurring. It is believed that with systematic effort this goal can be achieved. Accordingly, any such serious accident event normally comes as a surprise (as discussed by, for example, Turner and Pidgeon 1997); it is a black swan for those involved in the operation and management of the activity. Hence, one must be careful in describing the perspective taken when discussing whether an event is a black swan.

Formally, risk related to an activity may be defined, in its broadest sense and in line with the above reasoning, as (A, C, U), where A is the events occurring, C the related consequences (seen in relation to the values considered), and U the uncertainties (what will the consequences of the activity be?). In a risk assessment, risk is described by (A', C', Q, K), where A' and C' are the events and consequences specified in the assessment, Q the measure (in a wide sense) describing the uncertainties, and K the background knowledge that A', C' and Q are based on.

In an actual case, the A' events may not capture the A; we may experience some surprises relative to A'. For example, if we study the life of a young person, he or she may die of a disease not known today; the A' events do not cover the true A. For a black swan of type (b), A' are those events that we have identified in the risk assessment, and A is the occurrence of the actual event, which is a type of event known by others than those involved in the risk assessment. The event is a black swan of this type (unknown known) if A is not covered by A'. If A is not known at all, it is a black swan of type (a), an unknown unknown.

The third type of black swan events, (c), comprises those that are on the list of known events—A is captured by A' in the risk assessment, but the probability of occurrence of this A' is judged negligible and not believed to occur.

Paté-Cornell (2012) discusses the black swan metaphor and relates it to the "perfect storm" concept. This storm was a result of the combination of a storm that started over the United States, a cold front coming from the north, and the tail of a tropical storm originating in the south. All three meteorological features were known before and occur regularly, but the combination is rare. The crew of a fishing boat decided to face the storm, but they had not foreseen its strength. The storm strikes the boat, it capsizes and sinks; nobody survives (Paté-Cornell 2012).

This extreme storm is now used as a metaphor for a rare event that may occur, where we understand the underlying phenomena. The experts can calculate the probabilities of such events and the associated risks with a high degree of precision. They can make accurate predictions of future events, stating for example that in one in ten such situations the waves will be like this, and in one in 100 such cases the waves will become so big, etc. The probabilities are frequentist probabilities, characterising the variation in the phenomena, and they are known to a degree that is considered as certainty.

The black swan type of event (c) seems to be covered by the rare event of the perfect storm form. However, for this black swan type, we are in general in a situation where we cannot make this type of accurate prediction. We need to rely on subjective judgements where probability refers to a subjective or knowledge-based assignment of uncertainties and degrees of belief. When expressing that an event is judged to have negligible probability and not believed to occur, it is with reference to such a perspective.

If we have a situation with perfect information about the variation of the phenomena—we know the frequentist probability distribution (we are in the perfect storm situation)—one can argue that the occurrence of a low frequentist probability event should not come as a surprise. It is rare, but it is known with certainty that the event will occur sooner or later. Hence it is not a black swan (type c). However, one can also argue differently. Given the knowledge about the variation in the phenomena, it is considered so unlikely that the event will occur the next year, say, that it is not believed to occur. Hence it can be viewed as a black swan of type (c) if it in fact does occur. Again we see that whether the event is a black swan or not is in the eyes of the beholder.

In real situations, we cannot fully understand the variation. If we go back some years, we would not think of terrorism events as a contributing factor to the performance variation, and hence a black swan may occur even if the phenomenon is considered well understood. However, this type of black swans is not of type (c), but of type (a) or (b). The discussion here is linked to the distinction between common-cause variation and special-cause variation in the quality management field (Shewhart 1931, 1939; Deming 2000; Bergman 2009). The common-cause variation captures "normal" system variation, whereas the special causes are associated with the unusual variation and the surprises, the black swans; see Aven (2014).

We may also talk about "near-black swans," meaning surprises relative to one's knowledge/beliefs, where the event did not result in extreme consequences (for example as a result of the barriers working and avoiding the extreme outcomes). A black swan can occur as a result of a set of conditions and events, and a subset of these may generate a near-black swan.

How to confront the black swans

It is obviously not straightforward to assess and manage the black swan type of risk, and different approaches are recommended. Here are some of the most common approaches suggested:

- focus on signals and warnings;
- sensitivity to operations (principle of collective mindfulness linked to high-reliability organisations);
- adaptive risk analysis;
- robustness analysis;
- resilience analysis.

These approaches are discussed by, for example, Paté-Cornell (2012), Hollnagel et al. (2006), Weick and Sutcliffe (2007) and Aven (2014, 2015).

The essence of these approaches can be summarised as follows. We move forward in time and face some signals that make us reflect, adjust, and maybe stop. A signal in relation to risk is an event, a change in condition, etc., which provides information that something might happen. Think of a processing plant, where the pressure in a tank increases beyond normal or a valve begins to leak. We all know the signals coming from the body saying that now we have to take a break from work. If we ignore them, the consequences could be serious. It is important to be aware of these signals and react when they arrive because they make it possible to make adjustments and avoid negative outcomes. Unfortunately, it is not always easy to read the signals in the right way. The ability must be developed; one must find the balance between being vigilant and sensitive to signals, and being oversensitive, and respond to all sorts of changes. The right balance depends on a good understanding of risk—knowledge is the key. Some signals are not so important for safety and do not need immediate actions, and we will see this if the knowledge is strong. For example, when running we push the body forward; although one may feel tired at the start, it normally works fine. But there are symptoms and signals of a different kind: those that say that enough is enough, one must stop. Most people have experienced this and know how to distinguish between the different types of signals. Similarly, through good understanding of any system, through experience and learning, we acquire the ability to read the signals in the correct way. It is a continuous process. New and better understanding can nearly always be obtained.

Reading the signals, adapting, and adjusting, are essential features of these approaches. So is the ability to withstand the stressors that arise, being robust and resilient so that one can return to the normal state and even improve. These are all challenging tasks, as discussed in the above-cited references. In the following we will look more closely into two issues:

- how risk assessment can be improved to better address the black swan type of risk;
- how to extend current risk management frameworks to ensure that sufficient attention is devoted to the black swan risk.

Improved risk assessments

Black swan type of events, and near-black swan events occur. To confront them knowledge is the key, as highlighted above. Risk assessment is about producing this knowledge and hence we need to look into these assessments and see if we can improve them to better incorporate the black swan type of risk. Compared to current practice, many improvements can be made. Here is a checklist for those aspects to consider in a risk assessment to ensure that this type of risk is given adequate attention.

Checklist

1. Is there an overview of the assumptions made (linked to system, data, models, expert reviews, etc.)?
2. Has a risk assessment of the deviations from assumptions been conducted (an assumption deviation risk assessment)?
3. Have attempts been made to reduce the risk contributions from the assumptions that have the highest deviation risk?
4. Has the goodness of the models used been assessed? Have the model errors (difference between the correct value and the model's outcome) been found acceptable?
5. Has the strength of knowledge, on which the assigned probabilities are based, been assessed?
6. Is this strength included in the risk description?
7. Have attempts been made to strengthen the knowledge where it is not considered strong?
8. Have special efforts been made to uncover the black swans of the type unknown knowns?
9. Have special efforts been made to uncover any weaknesses or holes in the knowledge on which the analysis group has built their analysis?
10. Have special efforts been made to assess the validity of the judgements made where events are considered not to occur due to negligible probability?
11. Have people and expertise, who do not belong to the initial analysis group, been used to detect such conditions?
12. If the expected values of a quantity are specified, has the uncertainty of this quantity been assessed (for example, using a 90% uncertainty interval for this quantity)?

The assumption deviation risk method can be explained as follows, using the general risk terminology (C', Q, K) and letting D denote the deviation. The deviation risk is expressed as (ΔC', Q, K_D), where ΔC' is the change in the consequences (which includes D), and K_D

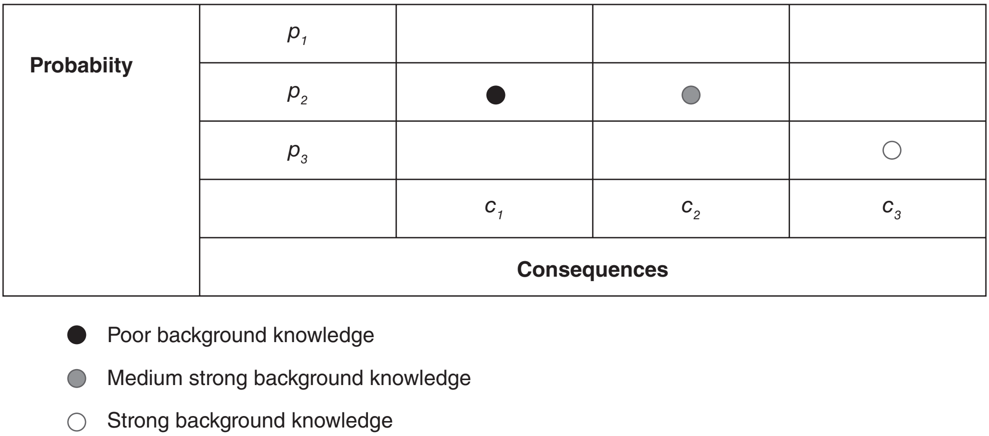


Figure 2.2 A risk matrix, based on specified consequences and probabilities, which incorporates the strength of knowledge. Here the c_is and p_is can alternatively represent intervals.

is the knowledge on which the ΔC' and Q are based. In practice, the method can be carried out by first providing a risk score on the basis of a judgment of D and related probabilities P (i.e. Q = P) (having in mind the implications for C'), and making an adjustment by a consideration of the strength of the knowledge K_D; see Aven (2013a) for the details.

The strength of knowledge could be represented as, for example, shown in Figure 2.2, in relation to a standard risk matrix. To give a score for the strength of knowledge, different approaches can be used as discussed in Aven (2013a). A simple one is just to provide a score based on an evaluation of the following type of criteria (Flage and Aven 2009):

1 the degree to which the assumptions made represent strong simplifications;
2 the degree to which data are not available or are unreliable;
3 the degree to which there is lack of agreement/consensus among experts;
4 the degree to which the phenomena involved are not well understood; models are non-existent or known/believed to give poor predictions.

An alternative and related approach is based on the work by Funtowicz and Ravetz (1990), (see also Kloprogge et al. 2005, 2011), linked to the NUSAP system. NUSAP was proposed to provide an analysis and diagnosis of uncertainty in science for policy by performing a critical appraisal of the knowledge base behind the relevant scientific information. The basic idea is to qualify quantities using the five qualifiers of the NUSAP acronym: numeral, unit, spread, assessment, and pedigree. Here our focus is on the last dimension, the pedigree. The pedigree approach is founded on a set of criteria to trace the origin of certain assumptions (hence, a "pedigree of knowledge") and qualify the potential value-laden character of these assumptions (Laes et al. 2011). The criteria used to discuss the assumptions are (van der Sluijs et al. 2005a, 2005b):

- *influence of situational limitations*: the degree to which the choice for the assumption can be influenced by situational limitations, such as limited availability of data, money, time, software, tools, hardware, and human resources;

- *plausibility*: the degree, mostly based on an (intuitive) assessment, through which the approximation created by the assumption is in accordance with "reality";
- *choice space*: the degree to which alternatives were available to choose from when making the assumption;
- *agreement among peers*: the degree to which the choice of peers is likely to coincide with the analyst's choice;
- *agreement among stakeholders*: the degree to which the choice of stakeholders is likely to coincide with the analyst's choice;
- *sensitivity to the view and interests of the analyst*: the degree to which the choice for the assumption may be influenced, consciously or unconsciously, by the view and interests of the analyst making the assumption;
- *influence on results*: in order to be able to pinpoint important value-laden assumptions in a calculation chain, it is important not only to assess the potential value-laden-ness of the assumptions, but also to analyse the influence of the assessment on outcomes of interest.

Each assumption is given a score between 0 and 4. The methodology is used to identify, prioritise, analyse, and discuss uncertainties in key assumptions. The general form of the pedigree as presented above has a typical societal safety/security scope and extends beyond the strength of knowledge considerations 1–4 above, in that, for example, views of external stakeholders are also incorporated.

Many types of risk assessment methods address the issue of what can happen, for example HazId, HAZOP, FMEA, fault tree and event tree analysis (Zio 2007; Meyer and Reniers 2013). One non-traditional method worth mentioning here is anticipatory failure determination (AFD) (Kaplan et al. 1999). It is a hazard/threat identification and assessment approach, which utilises I-TRIZ, a form of the Russian-developed Theory of Inventive Problem Solving. The AFD-TRIZ goes one step further compared to the traditional approaches and questions: "If I wanted to create this particular failure, how could I do it?" The power of the technique comes from the process of deliberately "inventing" failure events and scenarios (Masys 2012). A key task of the event/scenario identification is to challenge the assumptions, hypotheses, and explanations provided. A method often used for this purpose is *red teaming*, which serves as a devil's advocate, offering alternative interpretations and challenging established thinking (Masys 2012). In the following, an example of how a red-team analysis can be used to improve standard event/scenario identification is outlined (Veland and Aven 2015).

Adjusted event/scenario process using a red team

The assessment process has three main stages, involving two analyst teams, referred to as teams I and II. In *Stage 1*, analyst team I performs standard event/scenario identification (let us denote the events/scenarios by A_1'). This stage includes a self-evaluation of the analysis, where the focus is on the mental models (assumptions etc.) that could restrict the event/scenario's space.

In *Stage 2*, analyst team II challenges team I and their mental models, acting as a red team (the devil's advocate), and for example:

- argues against the mental models used by team I
- searches for unknown knowns.

A main purpose of the stage is to identify and assess potential surprises and black swans (surprising extreme events relative to one's beliefs/knowledge).

In the final *Stage 3*, the two analyst teams are to provide a joint list A_2.

In the case of unknown knowns, as some people - but not those involved in the risk assessment and the related management process – possess the knowledge, the analysis aims to achieve knowledge building and transfer of knowledge. To obtain such results, communication is essential. It could be communication between groups of people from different organisational units and between individuals. We may think of the Deepwater Horizon accident, where a worker did not alert others on the rig as pressure increased on the drilling pipe, a sign of a possible "kick" (Financial Post 2013). A kick is an entry of gas/fluid into the wellbore, which can set off a blowout. The worker had the knowledge (information), but this knowledge was not communicated to the right people.

Extension of current risk management frameworks

Current risk management frameworks are to a large extent probability-based and do not reflect the black swan type of risk as here described. To extend these frameworks, the following points, amongst others, need to be reflected:

- A sufficiently broad understanding of the risk concept, giving due attention to the knowledge dimension (such as the (C,U) risk perspective) is needed.
- Acceptable risk should not be determined by judgement about probability alone.
- Events may occur even if very low probabilities are assigned – the risk management needs to acknowledge this and have adequate means to meet this type of risk.
- The cautionary and precautionary principles constitute essential pillars of the risk management linked to such events (black swans).

If the risk management is built on a probability-based perspective, the black swan type of risk will not easily be given the attention it deserves, as the framework does not encourage considerations of the risk associated with the knowledge on which the probabilities are based.

In engineering contexts it is a common practice to use probabilistic criteria, like a 1×10^{-4} probability limit, to determine what an acceptable design is (Aven and Vinnem 2007). Such an approach cannot in general be justified as it ignores the degree of knowledge that supports the probability assignments, as discussed above. For a general discussion of such criteria, see Aven and Vinnem (2007). These authors state that such criteria must be used with care as they can easily lead to the wrong focus, meeting the criteria instead of finding the overall best arrangements and measures. However, for the practical execution of risk management activities, it is not difficult to see that some type of criteria may be useful in simplifying the decision-making process. To be able to meet the above critique related to the strength of knowledge supporting the probability assignments, an adjusted procedure has been suggested if such criteria are to be used (Aven 2013a):

- If risk is found acceptable according to probability with large margins, the risk is judged as acceptable unless the strength of knowledge is weak (in this case the probability-based approach should not be given much weight).

- If risk is found acceptable according to probability, and the strength of knowledge is strong, the risk is judged as acceptable.
- If risk is found acceptable according to probability with moderate or small margins, and the strength of knowledge is not strong, the risk is judged as unacceptable and measures are required to reduce risk.
- If risk is found unacceptable according to probability, the risk is judged as unacceptable and measures are required to reduce risk.

The approach relies on cautionary thinking. It generates a process that looks for measures to reduce risk and avoid the event occurring – despite the fact that the judged probability is very low. The cautionary principle states that, in the face of uncertainty, *caution* should be a ruling principle, for example by not starting an activity or by implementing measures to reduce risks and uncertainties (Aven and Renn 2010). The level of caution adopted has, of course, to be balanced against other concerns, for example costs, but to be cautious goes beyond balancing the expected benefit of risk reductions expressed in monetary terms against expected costs. The precautionary principle may be considered a special case of the cautionary principle, in that it applies in the face of *scientific uncertainties* (Aven 2011).

A number of safety and security measures are justified by reference to the cautionary principle. We implement robust design solutions to be able to meet deviations from normal conditions; we implement emergency preparedness measures even if the probability of their use is very small, etc. (Aven and Vinnem 2007). We see beyond the probabilities because we know that surprises can occur relative to our judgements. This is to be cautious. The risk related to a blowout is accepted when making decisions about the operation of, for example, an offshore oil and gas installation. The point being made here is that this risk cannot be made on probability judgements alone. The risk management systems need to have processes that scrutinise these judgements and their basis, and lift the considerations to a level that also takes into account the limitations of the analyses.

Conclusions

To confront possible black swans, we need to see beyond current risk assessment and management practices. It is a research challenge. In the present chapter we have looked into some areas that are considered important for meeting this challenge, but much more work has to be conducted to equip the risk community with well-proven approaches and methods to adequately assess and manage the black swan type of risk. The present chapter has reviewed some recent contributions to this end that focus on conceptualisation of the black swan risk and the type of frameworks that can be used for properly understanding, assessing, and managing this type of risk.

Knowledge and uncertainty are key concepts. Black swans can be seen as surprises in relation to someone's knowledge and beliefs. In the September 11 example, some people had the knowledge, others did not. In the Fukushima example, it was the judgements and probabilities that were essential, but they are based on data, information and arguments/opinions, so here too the issue is knowledge. We need new principles and methods to deal with this type of risk. The black swan type of research contributes to this end. This research lays the foundation for further development of methods and thereby helps in reducing the probability of (negative) black swans.

We need a risk thinking that allows for and encourages considerations and reinterpretations of the way risk is assessed at different stages of an activity, adequately reflecting the knowledge dimension. These are essential features of a management regime supporting continuous improvements. Current risk perspectives are considered to be less satisfactory for this purpose as the frameworks presume a stronger level of stability in the processes analysed.

References

Aven, T. (2011) On different types of uncertainties in the context of the precautionary principle. *Risk Analysis*, 31(10), 1515–1525. With discussion 1538–1542.

Aven, T. (2013a) Practical implications of the new risk perspectives. *Reliability Engineering and System Safety*, 115, 136–145.

Aven, T. (2013b) On the meaning of the black swan concept in a risk context. *Safety Science*, 57, 44–51.

Aven, T. (2014) *Risk, Surprises and Black Swans*. Routledge, New York.

Aven, T. (2015) Implications of Black Swans to the foundations and practice of risk assessment and management. *Reliability Engineering and System Safety*, 134, 83–91. Open Access.

Aven, T., Baraldi, P., Flage, R. and Zio, E. (2014) *Uncertainties in Risk Assessments*. Wiley, Chichester.

Aven, T. and Krohn, B.S. (2014) A new perspective on how to understand, assess and manage risk and the unforeseen. *Reliability Engineering and System Safety*, 121, 1–10.

Aven, T. and Renn, O. (2010) *Risk Management and Risk Governance*. Springer Verlag, London.

Aven, T. and Vinnem, J.E. (2007) *Risk Management*. Springer Verlag, Berlin.

Bergman, B. (2009) Conceptualistic pragmatism: a framework for Bayesian analysis? *IIE Transactions*, 41, 86–93.

Deming, W.E. (2000) *The New Economics*. 2nd edn. MIT-CAES, Cambridge, MA.

Financial Post (2013) Deepwater rig worker weeps as he admits he overlooked warning of blast that set off America's worst environmental disaster. http://business.financialpost.com/2013/03/14/halliburton-worker-weeps-as-he-admits-he-overlooked-warning-of-blast-that-set-off-americas-biggest-oil-spill-in-gulf/?__lsa=42e0-28bb. Accessed 15 January 2014.

Flage, R. and Aven, T. (2009) Expressing and communicating uncertainty in relation to quantitative risk analysis (QRA). *Reliability and Risk Analysis: Theory and Applications*, 2(13), 9–18.

Funtowicz, S.O. and Ravetz, J.R. (1990) *Uncertainty and Quality in Science for Policy*. Kluwer Academic Publishers, Dordrecht.

Gross, M. (2010) *Ignorance and Surprises*. MIT Press, London.

Hammond, P. (2009) Adapting to the entirely unpredictable: Black swans, fat tails, aberrant events, and hubristic models. *The University of Warwick Bulletin of the Economics Research Institute*, 2009/10(1), 2–3.

Hollnagel, E., Woods, D., and Leveson, N. (2006) *Resilience Engineering: Concepts and Precepts*. Ashgate, Aldershot, UK.

Kaplan, S., Visnepolschi, S., Zlotin, B., and Zusman, A. (1999) *New Tools for Failure and Risk Analysis: Anticipatory Failure Determination (AFD) and the Theory of Scenario Structuring*. Ideation International Inc., Southfield, MI.

Kloprogge, P., van der Sluijs, J., and Petersen, A. (2005) *A Method for the Analysis of Assumptions in Assessments*. Netherlands Environmental Assessment Agency (MNP), Bilthoven, the Netherlands.

Kloprogge, P., van der Sluijs, J.P., and Petersen, A.C. (2011) A method for the analysis of assumptions in model-based environmental assessments. *Environmental Modelling and Software*, 26, 289–301.

Laes, E., Meskens, G., and van der Sluijs, J.P. (2011) On the contribution of external cost calculations to energy system governance: The case of a potential large-scale nuclear accident. *Energy Policy*, 39, 5664–5673.

Lindley, D.V. (2008) The Black Swan: the impact of the highly improbable. Reviews. *Significance*, 5(1), 42.

Masys, A.J. (2012) Black swans to grey swans: revealing the uncertainty. *Disaster Prevention and Management*, 21(3), 320–335.

Meyer, T. and Reniers, G. (2013) *Engineering Risk Management*. De Gruyter Graduate, Berlin.

National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling (2011) *Deepwater: The Gulf Oil Disaster and the Future of Offshore Drilling*. Report to the President. US Government Printing Office, Washington, DC.

Paté-Cornell, M.E. (2012) On black swans and perfect storms: risk analysis and management when statistics are not enough. *Risk Analysis*, 32(11), 1823–1833.

Shewhart, W.A. (1931) *Economic Control of Quality of Manufactured Product*. Van Nostrand, New York.

Shewhart, W.A. (1939) *Statistical Method from the Viewpoint of Quality Control*. Dover Publications, Washington DC.

Smithells, R.W. and Newman, C.G.H. (1992) Recognition of thalidomide defects. *Journal of Medical Genetics*, 29(10), 716–723.

Taleb, N.N. (2007) *The Black Swan: The Impact of the Highly Improbable*. Penguin, London.

Taleb, N.N. (2012) *Anti Fragile*. Penguin, London.

Turner, B. and Pidgeon, N. (1997) *Man-Made Disasters*. 2nd edn. Butterworth-Heinemann, London.

van der Sluijs, J., Craye, M., Futowicz, S., Kloprogge, P., Ravetz, J. and Risbey, J. (2005a) Combining quantitative and qualitative measures of uncertainty in model-based environmental assessment. *Risk Analysis*, 25(2), 481–492.

van der Sluijs, J., Craye, M., Funtowicz, S., Kloprogge, P., Ravetz, J. and Risbey, J. (2005b) Experiences with the NUSAP system for multidimensional uncertainty assessment in model based foresight studies. *Water Science and Technology*, 52(6), 133–144.

Veland, H. and Aven, T. (2015) Improving the risk assessments of critical operations to better reflect uncertainties and the unforeseen. *Safety Science*, 79, 206–212.

Weick, K. and Sutcliffe, K. (2007) *Managing the Unexpected: Resilient Performance in an Age of Uncertainty*. Jossey Bass, CA.

Zio, E. (2007) *An Introduction to the Basics of Reliability and Risk Analysis*. World Scientific Publishing, Singapore.

Part 2

Managing the risks of extreme events

3 "As if their lives depended on it"

High performance in extreme environments

Ron Westrum

Introduction

One key to mission success in extreme environments is the use of relevant knowledge to provide optimal performance. Knowledge provides the tools for the Artic explorer and the high-performance rescue team alike to survive and prosper in ice and snow or high seas. Knowledge is likewise the key to "smart" weapons and tactics, using precise information to fine-tune the aggressive response. The knowledge must be located or generated. It must be embodied in equipment, training, and procedures. And most of all, it must be used in practice. These principles sound simple and obvious, but it is disturbing to see how often they are neglected.

Of course knowledge is important for virtually all missions, but in extreme environments, failure comes with high costs. People usually enter extreme environments because they must. It may be so because that is where the action is (e.g., combat), where the victims are (e.g., rescue), where the technology is highly complicated (cf. electric grids), or where the data must be collected (e.g., science in the Arctic). The presence of human beings thus seems non-negotiable. But the risks of human presence need to be managed. How can we do that?

Certain scholars seem to feel that the quest is hopeless and doomed to fail. This would be the view one receives from reading Charles Perrow's work on "normal accidents" (1984). According to Perrow, both the tempo and the complications of high technology frequently exceed human ability to follow, control, and make safe complex systems such as nuclear power, the transport of dangerous materials, and large-scale chemical processing. Furthermore, computers, complex coupling, and other features are certain to confuse and frustrate human ingenuity. One need only look to the Fukushima nuclear accident in Japan (March 11, 2012) to see how a witches' brew of natural disaster and human culpability can produce a catastrophe. An earthquake (9.0 Richter), a 45-foot high tsunami, and many faults built into the nuclear power system conspired to produce death, injury, and property destruction, as well as serious environmental pollution. Could it possibly have been avoided? (See Fackler 2012.)

For certain scholars, such accidents could be seen as symptoms of faulty organizations. The possibility, then, is that there are others that are non-faulty. The literature on "high reliability" (e.g., Weick 1987; Weick and Sutcliffe 2007) and "resilience" (Hollnagel et al. 2006), for instance, comes to such a conclusion. While "normal accidents" are common in third-world countries, first-world countries have developed organizations that mostly foresee and avoid accidents. These organizations, called high-reliability organizations, are able to manage dangerous and unpredictable systems because their organizational structure and culture has the "requisite variety" to respond to the exigent demands of the dangerous environment or the dangerous technology. Whether nuclear power plant, electric grid, oil platform, or crack medical team, the high-reliability team or organization knows what to anticipate, what to do, and how to recover (Westrum 2006). The combination of experience, skill, and sensible anticipation allows the humans to weave with the situation, and pull the chestnuts out of the fire, protecting the civilians as well as the practitioners from emerging dangers. Furthermore, organizations capable of high reliability are also typically more resilient; that is, able to survive threat and turmoil (Hollnagel et al. 2006). Yet, in spite of numerous case studies, and sometimes even "thick descriptions" of certain situations (e.g., Roe and Shulman 2008), there is still a lot to know about how reliability and resilience work in practice. This paper will explore the roles of information management and organizational learning in extreme environments, in search of further insights. We will look first at how information is gained, second at how it is stored, and third at how it is used in action.

Finding and getting the knowledge

Knowledge comes in various forms. It may be possessed by experts in universities or research institutes. It may be in the heads and hands of practitioners, who learn through experience. Or it may be something yet to be discovered. It may be down the hall, or around the world. The first step is getting it.

We might differentiate between theoretical knowledge, typically possessed by scientists, and practical knowledge, possessed by practitioners. Theoretical knowledge answers the question "Why does something happen?" whereas practitioner knowledge answers the question "How does something happen?" as well as the question "How do we make it happen or avoid it happening?" There are a number of disciplines, of which engineering is one example, that seek to bridge the gap between science and practice (cf. Barley 1996). Managing complex systems in extreme environments draws on both theory and practice.

In the Tokai-mura nuclear criticality accident, the knowledge was in the same building complex as the workers, but they didn't use it. The Tokai-mura accident took place in Japan on September 30, 1999. Workers were handling small amounts of nuclear fuel for a breeder reactor. The task involved mixing dissolved solutions of uranium. Already the workers were using an unapproved procedure, and with inadequate training. They proceeded to "innovate" their mixing operation further without seeking expert advice, and thus created an additional risk. In developing the new workaround, the workers did not consult the plant's engineers, yet they really did not understand in depth the physics of the practices they were going to use. Their workaround resulted in getting a critical mass of uranium in the mixing vessel. Three workers got dangerous levels of radiation; two died. Others also got lesser doses. Why did the workers not consult the engineers in the facility? Apparently, a major cause was lax supervision by management. Management had in fact encouraged the workers to save money and time by experimenting on their own. So the technicians moved into danger without grasping the risks they took (Westrum 2000).

Other teams, though, are more thorough in getting the information they need. So, for instance, in their studies of safety in aircraft carrier landings, the Berkeley group of high-reliability researchers (e.g., Rochlin et al. 1987; Weick and Roberts 1993) stressed that often the key to high performance was informal organizational learning. Such informal knowledge of the techniques used to land aircraft is gained through trial and error. It is passed through informal contacts (such as annual meetings of Navy bosuns), imprinted on new recruits, and constantly honed through practice, allowing highly skilled performance, with a resulting low risk of "crunches."

While the Berkeley school of thought stresses informal learning (including trial and error), there are other sources of learning. In major accidents, for instance, the technical community supports learning through the more formal processes that often follow such "tombstone" or "sentinel" events. Accident investigations are designed to pinpoint problems and suggest solutions. Thus, the resulting reports call for a close examination of extant ways of doing things, since current procedures have allowed a disaster to happen. Sentinel accidents found a herald in the British sociologist Barry Turner (1937–1995), whose book on *Man-Made Disasters* (1978, cf. Turner and Pidgeon 1997) was brilliant, but had little immediate impact. Then, in the 1980s, such disasters found their great advocate in the person of James Reason, professor of psychology at the University of Manchester. Reason's writings drew on the often lengthy reports of accidents; e.g., by judges (Reason, 1990). Reason soon became the leader of a large number of accident researchers. One of Reason's most important insights was to emphasize the role of unseen and insidious factors ("latent pathogens") in creating accidents, and to suggest that the removal of these factors was a most important activity. The investigations of sentinel events often bring latent pathogens into focus, allowing their removal, and providing lessons in avoiding them.

A perfect example of such major learning from sentinel events would be the investigations that followed the July 29, 1967 fire on the deck of the U.S.S. *Forrestal*, a large aircraft carrier (Stewart 2004). The fire, due to a Zuni rocket firing off on deck, was devastating, leading to a chain reaction, including detonation of bombs both on deck and on aircraft, as well as spillage and ignition of hard-to-quench aviation fuel. The first explosions wiped out most of the trained firefighters on the ship, causing the survivors to improvise measures that were not always successful. By the time the fire was put out, there had been 134 deaths, 161 further injuries, the loss of over two dozen aircraft and $72 million damage to the carrier itself—a major catastrophe.

The fire's aftermath led to two major reports, one by Rear Admiral Forsyth Massey on the fire itself, and a broader report by retired (four-star) Admiral James Russell on the problems of aircraft carrier fires generally. This second report was commissioned by, and had the strong backing of, the Chief of Naval Operations (CNO), Admiral Thomas Moorer. (The CNO is the highest operational position in the Navy.) These two reports suggested extensive equipment and training improvements that were contained in two long lists of recommendations. The major recommendation was that a better system for fighting fires, with better equipment, be used. For instance, one of the material improvements was to use better (i.e., neoprene) hoses, which were less likely to become tangled. The major human factors improvement was greater training in fighting fires. When a second major fire occurred on another carrier, the U.S.S. *Enterprise*, 18 months after the *Forrestal* fire (January 14, 1969), investigation showed for example that the training of the ship's crew in firefighting was far superior to that in the *Forrestal* fire. An excerpt from Rear Admiral Frederic A. Bardshar's report (February 11, 1969) states that

> The high state of training which existed aboard Enterprise produced the individual leadership at all levels which is necessary to an effective damage control organization... After each major explosion hose teams regrouped and resumed their efforts. When men fell, trained backup men took their place. In any event, the aggressive but

> controlled efforts of these fire fighting crews prevented the explosion of more 500 pound bombs which almost certainly would have occurred had the fires been allowed to burn unopposed.
>
> (cited in Stewart 2004, p. 72)

Nonetheless, 27 sailors died in the *Enterprise* fire, and another 344 were injured, 65 seriously. A report on the *Enterprise* fire was completed by Rear Admiral Bardshar, who noted which recommendations from the *Forrestal* fire had been implemented; many had, but others had not. Many other interesting features of the embodied learning from the two carrier fires can be gained from reading the Stewart (2004) thesis, but this brief outline suggests the value of such formal investigations and their sequels.

Another source of knowledge is personal experimentation. In Roald Amundsen's pioneering attempt to reach the North Pole, he took a number of measures to prepare himself to arrive there safely. He trained his body to endure hardship, he apprenticed with Eskimos to see how they endured the cold, and he noted the dynamics of using sled dogs and the value of eating dolphin meat. In deployment to the South Pole (he had changed his mind when someone else reached the North Pole first), Amundsen lowered his risk by prepositioning supplies and marking his route with black flags in precise increments of distance. In contrast to another South Pole explorer, Robert Scott, Amundsen bought more supplies than he thought he would need, and brought four thermometers compared to Scott's one. Amundsen reached the South Pole and returned, while Scott died in the frozen wastes, ten miles from a supply depot he never reached (Collins and Hansen 2011, p. 17). The experimentation that Amundsen carried out reduced his risk level, and enabled him to succeed where Scott failed.

The embodiment of knowledge

Getting the knowledge is not enough. The knowledge must be embodied, and thus made "permanent" in the organization. It may be embodied in things (via design) or embodied in people (human capital). Of course, in either case it may also eventually be forgotten or otherwise lost (Martin de Holan and Phillips 2011).

In the case of the two carrier fires, the "learnings" were contained in the three reports—two from *Forrestal*, one from *Enterprise*. The possibility that these reports would lie unread was present; hence the emphasis on getting the material turned into hardware and software, and the requirements in training turned into better training practices and classroom capacity. The Navy also requested and for many years got three-month periodic updates on the progress of the recommendations of the Russell report. These kept track of which of the recommendations had resulted in changes. The support of Admiral Moorer was key in getting many of the changes implemented. In the Navy, it is routine to rotate the officer who is CNO; Admiral Moorer, who had overseen the reports and changes stemming from the two fires, left office on 1 July, 1970, and was replaced by Admiral Elmo Zumwalt. It is notable that the updates on the changes recommended by the fire investigation continued to arrive until 1972, however.

Learning from experience can be embodied in equipment. For instance, a prime example of this is the design of the Grumman F6F aircraft, the Hellcat, probably the key plane used by the U.S. Navy in WWII. The Hellcat's design was begun early in the 1940s, but information from pilots and combat was continually fed into the ongoing design until the aircraft was released in 1943 (Anderson 1996). It is said that Leroy

("Roy") Grumman, head of Grumman Aircraft, threw a big party for the fleet pilots after the battle of Midway (June 4–7, 1942) at the Royal Hawaiian Hotel. Everything was laid on: liquor, food, and attractive women. During the party, Grumman circulated among the pilots, trying to discover which features the pilots wanted in the standard carrier-borne fighter, at that time the F4F Wildcat. Did they want, for instance, a plane that could turn and maneuver as fast as the Japanese "Zero"? Apparently not. What they said they wanted, instead, was better machine guns and longer-range gas tanks. Soon, they would get those things in the Grumman F-6F Hellcat. Whether the story is true or apocryphal, the basic point is that Grumman was highly sensitive to the pilots—especially those who had flown the predecessor F4F Wildcat fighter. It embodied the lessons of previous combat in shaping the Hellcat.

Another input to the design of equipment takes into account "What might go wrong?" Dealing with high-risk situations can be fatal unless what I have called "requisite imagination" is used to grasp and mitigate the risks. It means that you try to imagine anything that could go wrong, and develop a solution for dealing with it—in advance (Adamski and Westrum 2003; Westrum 2009). For instance, in designing the moon landers, the Grumman company had to take into consideration every likely scenario for failure of the lander's equipment, and as much as possible, design a fix (Kelly 2001). The lander, to take one aspect, might have to be used as a lifeboat if it became impossible to operate the command module, so the designers took this potential role into account; in fact, during the blowout crisis on *Apollo 13*, the lander was used as a lifeboat, since it took much less energy than the command module (Westrum 2009).

Still another way that knowledge can embodied is via a checklist (Gawande 2007, 2012). The checklist is a codification, in temporal order, of what to do with technology, and in what order. Gawande gives the example of a checklist with perhaps a thousand items being used to revive a three-year-old girl who had fallen into an icy alpine lake. When pulled out of the lake, she appeared to be nearly dead, but with careful treatment she eventually recovered all of her faculties. Gawande comments, "The degree of difficulty in any one of these steps is substantial. Then you must add the difficulties of orchestrating them in the right sequence, with nothing dropped, leaving some room for improvisation, but not too much" (Gawande 2007, n.p.). Another vital checklist, with only five steps, has been used to greatly reduce lethal central-line infections in hospitals (Pronovost 2010). Since hospital infections kill over 100,000 patients a year, this is highly significant.

Amazingly enough, organizations may sometimes save information, but in a form difficult to access. Many organizations that painstakingly pursue a "lessons learned" process will follow up by putting the information, quite literally, on the shelf. Many of my experienced students noted that even if their own organizations saved stories of previous examples, they failed to use the information the next time the problem arose. This was for two reasons. First, it was difficult to search and to access the lessons-learned database; second, there was seemingly never enough time, once development of a scenario started, to consult the database. The lucky developments resulted in the organization learning the same lesson a second time; the unlucky ones never knew what hit them. Retrieving saved "lessons learned" is an art in itself.

Learning all the specifics of a particular situation, and above all, learning to successfully use the equipment one has, is often key to high performance. Learning to use the equipment doesn't sound like a particularly clever principle, but it is much more common with high-performance teams than with middling performers. While Lord Nelson was quoted as saying that "No captain can do very wrong who lays his ship alongside that of the enemy," he was far from believing that simply closing with the enemy was enough. Simply showing up to fight was far from sufficient. Learning how to use the equipment most effectively was vital to the British Navy's—and especially Nelson's—success in the battles against Napoleon's fleet; see the comments of Toll (2007) regarding rate of fire. Nelson spent many hours with his captains discussing tactics, making sure they understood his principles of flexible fleet movement. Using unexpected maneuvers, Nelson brought a tactical flexibility to the former practice of the ships staying in a line. This tactical flexibility allowed Nelson success in the Battle of the Nile and similar battles (Mostert 2008).

As was the case with Nelson's captains, sometimes the knowledge is embodied in people. It may be transmitted through training, brought in via personnel transfers, or even imbued through on-the-job coaching.

In 1982, Argentina decided to take possession of the nearby Falkland Islands. The Falklands (also known as the Malvinas) were a British colony, located far down in the South Atlantic. Whatever the merit of the Argentinian claim to the islands, the UK under Margaret Thatcher was not about to let the islands fall into the hands of Argentina. The British sent a small fleet of warships, including originally two aircraft carriers and later one more. The two carriers they sent originally were both small, but carried the versatile Harrier jet fighter.

The two air squadrons operating from the Hermes and Invincible were the 800 and the 801. They had different attitudes toward preparation for combat, and especially the degree to which the two squadrons learned to depend on their "blue fox" aerial radar. Whereas the 801 had carefully explored the strengths and weaknesses of the radar, the 800 did not have this careful acquaintance and therefore thought that the radar was basically undependable. As a result, the 801 was able to successfully perform operations that the 800 could not. In his excellent autobiographical book on the conflict, Commander Sharkey Ward describes the many ways in which the 801's embrace of the radar empowered them, in contrast to the 800 squadron, which treated the radars with suspicion (Ward 1993).

Certainly, the prime example of this principle of "getting to know the equipment" was in the research and activities of Navy Captain Frank Ault, most famous for his role in the creation of Top Gun aviation training center in Miramar, California. During the Vietnam War, Ault had been the executive officer on an aircraft carrier, and found disturbing the low "exchange rate" in shoot-downs of American fighter pilots versus North Vietnamese pilots. In Korea, the Americans had had a much higher exchange rate. Ault thought that American pilots were getting shot down more often than they needed to. He wrote several letters to Washington, but nothing seemed to happen—until, that is, when one day he showed up in person in Washington. He soon found himself in the presence of two senior admirals—Rear Admiral Robert Townsend and Vice Admiral Thomas Tom Connolly. The two quickly put him in charge of a project to fix the problem:

> To my surprise I found that I had talked myself right into the hole that Townsend and Connolly had already dug and readied for me. I was given the "simple" mandate to find out why the navy was not shooting down more MiG's and to recommend what should be done about it. I was given *carte blanche* to pick anybody I wanted to assist me and to conduct the study in any manner I saw fit. As for desired end results,

> I was told "an increase in combat kills of not less than three times better than performance to date, and as soon as possible."
>
> (Ault 1989, p. 36)

Captain Ault quickly put together an interdisciplinary team of experts in five different areas, and giving himself a 90-day deadline, proceeded to use a variety of resources to research the situation. He compiled an action list of what needed to be done with missiles, from design to combat training. Many of these action items turned into actions. One example of the items on the list was a "graduate school for fighter weapons." In the months that followed, this vision became Top Gun, the Miramar "Fightertown," where pilots were taught how to dogfight with the enemy. The Phantom F-4 fighter in particular was carefully studied to optimize its use and tactics. When the pilots trained at Top Gun rotated to Vietnam, the result was a much improved exchange rate against the North Vietnamese air force (Westrum 1999, p. 217; Wilcox 1990). While Top Gun was important to the Vietnam conflict, it also became a permanent addition to the Navy training program—yet as we will see, while in one case there may be rapid assimilation and use of answers, nothing guarantees this outcome.

Using knowledge

High performance requires that the best knowledge be used. Often, it is not. Two business-oriented authors have written a best-selling book on what they call "the knowing-doing gap" (Pfeffer and Sutton 2000). Their approach is clearly germane to our subject. In many cases, the knowledge is present, but not acted upon.

We can see this failure to follow the best advice in the death of the top Formula One driver Ayrton Senna. It was May 1, 1984, during the Imola Race in Italy. Two serious crashes had already taken place. One driver, Rubens Barrichello, had been seriously injured on Friday during practice. Then, on Saturday, a crash put Roland Ratzenberger into intensive care; he soon died. Doctor Sid Watkins saw that Ayrton Senna was clearly upset. Watkins, the Formula One Association's neurosurgeon and one of Senna's close friends, did not want Senna to race. He said, "Ayrton, why don't you withdraw from racing tomorrow. I don't think you should do it. In fact, why don't you give it up altogether? What else do you need? You have been World Champion three times, you are obviously the quickest driver. Give it up and let's go fishing ... I don't think the risk is worth continuing—pack it in" (Watkins 1996, p. 8). That day, Senna said to Watkins, "Sid, there are things over which we have no control. I cannot quit. I have to go on." He did race, and had a fatal accident.

Watkins had a long and distinguished career as a neurosurgeon, but his passion was Formula One racing. Watkins had saved many drivers from injury and death by advising them on when they should and should not race. Getting Watkins involved with Formula One racing was an excellent move. Over the years, Watkins would do a great deal to make Formula One safer, but he could not save his close friend Senna from a fatal crash.

More often, teams manage to learn from experience. In contrast to the Imola accident, we have the following scenario. 1958 found Air Force Colonel Charles ("Chuck") Yeager preparing for a deployment to Europe of his (supersonic) F-100 fighter squadron, with 18 fighter aircraft. Two in-flight refuelings over the Atlantic for all the F-100s were involved. A fighter pilot who crashes in the Atlantic has few chances of survival, even if he can eject from the aircraft, so Yeager, who had been an aircraft mechanic himself, insisted on his pilots listening to their mechanics.

> Maintenance was the heart of a squadron, and our crews had everything they needed to keep everything in top shape. I just told the crew chiefs, "You guys are in charge. When you tell a pilot that his airplane is ready, that's all he needs to know. So, you'd better make damned sure you know what you're talking about."
>
> I never applied pressure to keep all our airplanes in the air; if two or three were being serviced, we just lived with an inconvenience, rather than risking our lives with aircraft slap-dashed onto the flight line. I wouldn't allow an officer-pilot to countermand a crew chief-sergeant's decision about grounding an unsafe airplane. A pilot faced with not flying wasn't always the best judge about the risks he was willing to take to get his wheels off the ground. And it paid off. My pilots flew confident, knowing that their equipment was safe.
>
> (Yeager and Janos 1985, p. 315)

Speaking of using the best knowledge, it is remarkable that quite often amateur pilots ignore the state of the weather. It is well known that doctors are among the most accident-prone private aircraft pilots, because they seem to believe that they can control the weather. An older study found that physicians in 1964–1965 were responsible for four times the accidents that their prevalence in the private pilot world would suggest (Mohler et al. 1966; see also Benson 2011). Physicians are not the only pilots that ignore weather warnings. Such ignored warnings also figured in the loss of the airplane flown by John Kennedy, Jr.

Pilots are not alone when arrogance suggests that good advice can be ignored. In Collins and Hansen's brilliant book *Great by Choice* (2011, pp.100–103), the authors relate the circumstances that led to one of the worst accidents on Mount Everest, where a party led by guide Robert ("Rob") Hall in 1996 was caught in a storm and eight people died. Hall had proceeded upward in the face of storm warnings that led a more prudent climber, David Breashears, to turn around in his tracks and take his IMAX camera-laden party down the mountain. The two parties in fact passed each other, Breashears' retreating downward and Hall's party ascending. Hall's party had eight deaths, whereas Breashears' group lived to go on to make a second, successful try, ending in an excellent photo shoot.

The value of careful training is underlined by early experiences with flight control in the U.S. National Aeronautics and Space Administration (NASA). Certainly, during the Apollo spacecraft program, the formation of the flight-control team was an excellent example of the creation of a skilled group that together could and did perform miracles (Kraft 2001). Eugene ("Gene") Kranz's "conglomerate brain" was able to formulate the answers needed to provide direction to operational problems in as few as 20 seconds (Chaikin 1995).

Following the explosion on *Apollo 13* on April 13, 1970, NASA flight control devoted its whole energies to the problem. During the crisis with *Apollo 13*, every available measure was taken to insure that the astronauts returned. Kranz's attitude that "failure is not an option" led the team to exceptional lengths, trying to find something, anything that would get the astronauts safely back to earth (Kranz 2000), and they succeeded.

Perhaps this attitude had taken root after the fire (January 27, 1967) in the space capsule for *Apollo I*. The fire caused the death of three astronauts: Gus Grissom, Ed White, and

Roger Chaffee. This fire had a tremendous impact on attitudes about safety at NASA. After the fire, Kranz had given the flight controllers a speech, as follows:

> From this day forward, Flight Control will be known by two words, "Tough" and "Competent." *Tough* means we are forever accountable for what we do or what we fail to do. We will never again compromise our responsibilities. Every time we walk into Mission Control we will know what we stand for. *Competent* means we will never take anything for granted. We will never be found short in our knowledge and in our skills. Mission Control will be perfect. When you leave this meeting today you will go to your office and the first thing you will do there is to write "Tough and Competent" on your blackboards. It will *never* be erased. Each day when you enter the room these words will remind you of the price paid by Grissom, White, and Chaffee. These words are the price of admission to Mission Control.
>
> (Kranz 2000, p. 204)

Later, Kranz would comment on *Apollo 13*,

> What we could not accomplish through technology, or procedures and operating manuals, we might be able to manage by drawing on a priceless fund of experience, accumulated over almost a decade of sending men into places far beyond the envelope of earth's protective, nurturing atmosphere. All we had to work with was time and experience. The term we used was "workaround"—options, other ways of doing things, solutions to problems that weren't found in manuals and schematics. These three astronauts were beyond our physical reach. But not beyond the reach of human imagination, inventiveness and a creed that we all lived by: "Failure is not an option."
>
> (Kranz 2000, p. 12)

The last Apollo mission would take place in December 1972—yet people leave and cultures change. However energized the Apollo teams might have been by the fire, a different culture seems to have settled in at NASA during the space-shuttle era. In 1986, the shuttle *Challenger* exploded when the booster seals were bypassed, allowing ignition of the fuel tank, a problem that had been spotted, but then ignored. Then, in the year 2003, we find the managers of the shuttle *Columbia* STS-107 in what seemed to be a repeat of the *Challenger* accident (Mahler and Casamayou 2009).

The *Columbia* shuttle was in potential danger; a "foam strike" on the wing root had created an unknown amount of damage. How much damage was there, and how could it be fixed or escaped? Structural engineer Rodney Rocha badly wanted to find out how much damage there was, so he could determine what had to be done—but in contrast to *Apollo 13*, Rocha's efforts to clarify the situation were blocked at every turn by managers who seemed to be mostly unresponsive to his concerns about getting more information. The managers had given up on trying to save the astronauts aboard. If something was found to be wrong with the flight, they felt there was nothing they could do. Fatalism seemed to have replaced extreme effort as the order of the day (Cabbage and Harwood 2004). In the end, Rocha did not get the pictures from the Air Force that he sought, he never got the information he needed, and the reentry of the *Columbia* into the atmosphere led to its destruction. There is a huge contrast here between the "business as usual" approach of Columbia's mission-management team and the attitude taken by NASA to save *Apollo 13*. Whereas Kranz seemed willing to move heaven and earth to save

Apollo 13's astronauts, the attitude of the Columbia mission-management team seemed almost like sleepwalking. It is difficult not to believe that this attitude contributed to the loss of the *Columbia* and its astronaut crew.

Discussion

How, then, are we to make sense of these different behaviors and performances? Why do some organizations seem to learn, whereas others do not? Is there really a kind of "learning organization," as Senge (1990) has suggested? Personally, I do not find Senge's theoretical insights strong. On the other hand, I have heard from practitioners that organizations that have adopted his methods have thrived. How can this be? Maybe the important thing is the stance or attitude of the organization's personnel thinking of themselves as a learning organization. I am not sure; for those seeking a broader review, I suggest reading Argote (2013).

On the other hand, my own orientation suggests a similar line of thought. I believe that organizational culture provides the important bridge from insight to application. In my own schema (see Figure 3.1), I have suggested that organizational cultures form a continuum from pathological to generative in how they process information (Westrum 2014). Whereas organizations on the pathological end tend to see information as a personal resource, those on the generative end tend to see it as a resource for the organization as a whole. Accordingly, information flows better toward the generative end of the Westrum continuum. Generative organizations, embodying higher cooperation and trust, encourage the sharing and transfer of information. So these organizations tend to be better able to incorporate new information.

For instance, I believe that during the Apollo program, the ability to listen to one generation after another of John Houbolt's ideas on a lunar-orbit lander depended importantly on the generative culture of that organization in the 1960s (Hansen, 2013). Without that ability to listen, the lunar-orbit solution would not have been adopted. As both of the major shuttle accidents suggest, however, NASA's culture had changed during the shuttle era. Materials on the *Columbia* tragedy (e.g., Cabbage and Harwood 2004) show strong evidence of an extreme bureaucratic culture, somewhere in the middle between pathological and generative. Information was bottled up, key information was not available or was not shared, and cooperation during the crisis was lacking. The culture had radically changed from the days of Apollo.

In the medical realm, there is a great deal of evidence that generative cultures are closely associated with a variety of good practices and outcomes (West et al. 2014). These

Pathological	*Bureaucratic*	*Generative*
Emphasis: promote *the chiefs* Information is hidden Messengers are shot Responsibility is shirked Bridging is discouraged Failure is covered up New ideas are crushed	*Emphasis*: protect *the department* Information may be ignored Messengers are tolerated Responsibility is defined Bridging is allowed Failure is dealt with fairly New ideas create problems	*Emphasis*: accomplish *the mission* Information is actively sought Messengers are trained Responsibility is shared Bridging is rewarded Failure causes inquiry New ideas are welcomed

Figure 3.1 How organizations process information.

include encouragement of worker involvement, a "level playing field," and inducements to "speak up."

Similarly, when there is a "political" (read "pathological") culture of accident investigation, the pressure to distort the facts is strong. It is well known that this has happened several times in major accident investigations:

1 the 1979 Mt. Erebus, New Zealand, crash (Vette 1984);
2 the 1994 RAF Chinook crash on Mull of Kintyre (Computer Weekly 1999);
3 the 1987 ATR-42-300 crash at Mt. Crezzo, Italian Alps (Fredrick 1996);
4 the 2006 RAF Nimrod crash in Afghanistan (Haddon-Cave 2009).

Good information flow is critical to the learning process. When the wrong cause of an accident is indicated, the chance to learn the correct lesson is foregone. To take only one example, key facts relevant to the 1987 Mt. Crezzo crash were suppressed by the Direction Générale de l'Aviation Civile in France (National Transportation Safety Board 1997). This was information that showed a design problem in the aircraft that had crashed, glossed over presumably for commercial reasons. The information omitted from the crash report may in turn have contributed to a second crash in 1994 at Roselawn, Illinois. Four crew members and 64 passengers died in this second crash (Fredrick 1996).

In the other three accidents as well, there was an attempt to blame the pilots for the crashes, whereas in fact other unseen factors were significant. In the Mt. Erebus case, the flight manager (computer) had been programmed incorrectly. In the Chinook and Nimrod cases, problems with software and maintenance, respectively, were strongly implicated. Only with further investigation did these causes come to light. Thus, the culture of investigation needs attention.

Conclusions

Use of appropriate knowledge is often a key factor in mission success and human survival in extreme environments. The exploration I have done here is merely a sampling from a much larger and richer universe. Getting knowledge, embodying it, and using it are all specific problems. Where do people (and organizations) find the information they need to get their jobs done? If they don't get it, what happens? Even if they have it, how do they then embed, or hold on to the knowledge they need? What happens when, through design or inadvertence, that knowledge is lost (cf. Franzen 1994)? What makes organizations able to use the knowledge that they have? Why do organizations, having established a successful pattern, do something else?

Some years back, it was found by the Veterans Administration (VA) that its doctors often failed to do routine procedures that saved lives. After a heart attack, there were nine procedures that much improved the patient's chances of survival, but many VA hospitals didn't do them. These procedures were important and were known to work well, to produce positive results. If you used these drugs, the mortality went down 40 percent! Why didn't they use them, then? The answer, it would appear, was that the doctors simply forgot them (Kolata 2004). Since the information wasn't embedded in a checklist, it didn't have compelling power over physicians, who relied instead on their all-too-faulty memories.

One would think that in life-and-death situations like heart disease, people would not take chances. One would think that, when one is handling nuclear isotopes, people would check with engineers about potential hazards, but the technicians didn't do that in the Tokai-mura accident. Nor did the technicians polishing the mirror in the Hubble Space Telescope check with the engineers before they inserted shims under the mirror while they were polishing it. The images taken by the main camera came back distorted. The Hubble's problems led to a second space-shuttle mission, to fix the telescope the technicians had screwed up.

Many of us have had the experience of taking home a consumer product that requires some assembly. Sometimes we fail to read the instructions, because it seems "obvious" how it is going to go together. After it doesn't quite come together, *then* we decide to read the instructions. A consumer product is one thing—but one would think that in extreme environments there would be more planning and "double-checking" to make sure things go according to plan. Surprisingly, sometimes there isn't.

By contrast, major accidents and disasters typically claim our attention. We know something is seriously wrong; we empanel a committee to find out what it is. The committee investigates, it reports out, changes are made. We are not going to have the same accident again, are we? However, having lost the shuttle *Challenger*, NASA proceeded to lose *Columbia*. Sometimes we don't learn, or don't care, or cannot afford to do the right thing. When Hurricane Katrina hit New Orleans, every police commander had a disaster-recovery plan in his or her office. The problem was, no one had read it, so recovery from the hurricane was developed largely from scratch. The knowledge was there, but no one had read the plan, and no one had practiced it.

Charles Franklin "Boss" Kettering (of General Motors) once said that "getting a new idea into a factory is the greatest endurance contest in the world." Well, how about getting an organization to learn a lesson?

References

Adamski, A., and Westrum, R., 2003. "Requisite Imagination: The Fine Art of Anticipating What Might Go Wrong," in Erik Hollnagel, ed., *Handbook of Cognitive Task Design*. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 193–222.

Anderson, D.A., 1996. "The Hellcat," in J. Ethell et al., eds, *The Great Book of World War II Airplanes*. New York: Crescent, pp. 157–212.

Argote, L., 2013. *Organizational Learning, Creating, Retaining, and Transferring Knowledge*. 2nd edn. New York: Springer.

Ault, F., 1989. Interview with Ron Westrum.

Barley, S.R., 1996. *The New World of Work*. London: C.D. Howe Institute.

Benson, G., 2011. "Gene's Top Ten List of Pilot Killers." www.genebenson.com/accident_prevention/top_ten/top_ten.htm. Accessed on July 15, 2015.

Cabbage, M., and Harwood, W., 2004. *Comm Check…: The Final Flight of Shuttle Columbia*. New York: Free Press.

Chaikin, A., 1995. *Man on the Moon: The Voyages of the Apollo Astronauts*. New York: Penguin.

Collins, J., and Hansen, M.T., 2011. *Great by Choice*. New York: Harper Collins.

Computer Weekly, 1999. *RAF Justice: How the Royal Air Force Blamed Two Dead Pilots and Covered up Problems with the Chinook's Computer System FADEC*.

Fackler, M., 2012. "Nuclear Disaster in Japan Was Avoidable, Critics Contend." *The New York Times*, March 10, pp. A4, A7.

Franzen, J., 1994. "Lost in the Mail." *The New Yorker*, October 24, pp. 62–77.

Fredrick, S.A., 1996. *Unheeded Warning: The Inside Story of American Eagle Flight 4184*. New York: McGraw-Hill.

Gawande, A., 2007. "The Checklist." *New Yorker*, December 10. http://www.newyorker.com/magazine/2007/12/10/the-checklist. Accessed on July 8, 2017.

Gawande, A., 2012. *The Checklist Manifesto: How to Get Things Right*. New York: Metropolitan Books.

Haddon-Cave, C., 2009. *The Nimrod Review: An Independent Review into the Broader Issues Surrounding the Loss of the RAF Nimrod MR2 Aircraft XV230 in Afghanistan in 2006*. London: H.M. Stationery Office.

Hansen, J.R., 2013. *Enchanted Rendezvous: John C. Houbolt and the Genesis of the Lunar Orbit Rendezvous Concept*. NASA Aerospace History Monograph #4. Amazon: Createspace, Nov. 2.

Kelly, T.J., 2001. *Moon Lander: How We Developed the Apollo Lunar Module*. Washington, DC: Smithsonian Institution Press.

Kolata, G., 2004. "Program Coaxes Hospitals to See Treatments under Their Noses," *The New York Times*, December 25, pp. A1, C6.

Kraft, C., 2001. *Flight: My Life in Mission Control*. New York: Dutton Adult.

Hollnagel, E., Woods, D., and Leveson, N., eds, 2006. *Resilience Engineering: Concepts and Precepts*. Aldershot, UK: Ashgate.

Kranz, E., 2000. *Failure is not an Option: Mission Control from Mercury to Apollo 13*. New York: Simon and Schuster.

Mahler, J.G., and Casamayou, M.H., 2009. *Organizational Learning at NASA: The Challenger and Columbia Accidents*. Washington, DC: Georgetown University Press.

Martin de Holan, P., and Phillips, N., 2011. "Organizational forgetting," in Easterby-Smith, M., and Lyles, M.A., eds, *Handbook of Organizational Learning and Knowledge Management*. New York: Wiley, pp. 433–451.

Mohler, S.R., Freud, S., Veregge, J., and Umberger, E., 1966. *Physician Flight Accidents*, Washington, DC: Federal Aviation Administration, Office of Aviation Medicine, September.

Mostert, N., 2008. The *Line Upon a Wind: An Intimate History of the Greatest War Fought at Sea Under Sail, 1793–1815*. New York: W.W. Norton.

National Transportation Safety Board, 1997. *Symposium on Corporate Culture and Transportation Safety*. April 24. Arlington, VA: National Transportation Safety Board.

Perrow, C., 1984. *Normal Accidents: Living with High-Risk Technologies*. New York: Basic Books.

Pfeffer, J., and Sutton, 2000. *The Knowing-Doing Gap*. Cambridge, MA: Harvard Business School Press.

Pronovost, D., 2010. *Safe Patients, Smart Hospitals*. New York: Penguin.

Reason, J., 1990. *Human Error*. New York: Cambridge University Press.

Rochlin, G., Laporte, T., and Roberts, K., 1987. "The Self-designing High Reliability Organization: Aircraft Carrier Flight Operations at Sea." *Naval War College Review*, 40(4), 76–90.

Roe, E., and Shulman, P., 2008. *High Reliability Management: Operating on the Edge*. Palo Alto: Stanford University Press.

Senge, P., 1990. *The Fifth Discipline: Strategy and Tools for Building a Learning Organization*. New York: Random House.

Stewart, H.P., 2004. *The Impact of the USS Forrestal's 1967 Fire on United States Navy Shipboard Damage Control*. Master of Military Science and Art thesis (History), Fort Leavenworth, Kansas.

Toll, I., 2007. *Six Frigates: How Piracy, War, and British Supremacy at Sea Gave Birth to the World's Most Powerful Navy*. New York: Penguin.

Turner, B., and Pidgeon, N., 1997. *Man-Made Disasters*. 2nd edn. London: Butterworth and Heinemann.

Vette, G., 1984. *Impact Erebus*. New York: Sheridan House.

Ward, S., 1993. *Sea Harrier over the Falklands*. London: Orion Press.

Watkins, S., 1996. *Life at the Limit: Triumph and Tragedy in Formula One*. London: Macmillan.

Weick, K., 1987. "Organizational Culture as a Source of High Reliability." *California Management Review*, 29(2), 112–127.

Weick, K., and Roberts, K., 1993. "Heedful Interrelating on Flight Decks." *Administrative Science Quarterly*, 38(3), 357–381.

Weick, K., and Sutcliffe, K.M., 2007. *Managing the Unexpected: Resilient Performance in an Age of Uncertainty*. 2nd edn. New York: John Wiley.

West, M.A., Topakas, A., and Dawson, J.F., 2014. "Climate and Culture for Health Care Performance," in Schneider, B., and Barbera, K., eds, *Oxford Handbook of Organizational Climate and Culture*. New York: Oxford University Press, pp. 335–359.

Westrum, R., 1999. *Sidewinder: Creative Missile Development at China Lake*. Annapolis, MD: Naval Institute Press.

Westrum, R., 2000. "Safety Planning and Safety Culture in the JCO Criticality Accident," *Cognitive Technology at Work*, 2, 240–241.

Westrum, R., 2006. "A Typology of Resilience Situations," in Hollnagel, E., Woods, D. D. and Leveson, N.C., eds, *Resilience Engineering: Concepts and Precepts*. Aldershot, UK: Ashgate, pp. 55-66.

Westrum, R., 2009. "Two Modes of Resilience," in Chris Nemeth, ed., *Perspectives in Resilience Engineering*, Vol. 2. Aldershot: Ashgate, pp. 55–66.

Westrum, R., 2014. "The Study of Information Flow: A Personal Journey. *Safety Science*, 67, 58–63.

Wilcox, R.K., 1990. *Scream of Eagles: Top Gun and the U.S. Air Victory in Vietnam*. New York, Wiley.

Yeager, C., and Janos, L., 1985. *Yeager: An Autobiography*. New York: Bantam.

4 Decision making on trial

The extreme situation at Fukushima Dai Ichi

Sébastien Travadel, Christophe Martin, and Franck Guarnieri

Introduction

The "decision" as "the end of deliberations about a voluntary act resulting in the choice of an action,"[1] emerges from an experiential background that is taken for granted, while at the same time it is the necessary substrate for action. On the basis of this definition, intention requires the subject to face their responsibilities; the evaluation of their choice highlights to the subject and to others, through its consequences, the arguments that underlie methods, processes, and other artifices of analysis that are used in the deliberative stage to optimize the decision. The observation of organizations, however, leads to the attenuation of the scope of this decision model, an emphasis on the influence of networks of actors, and the succession, sometimes more temporal than causal, of decisions, even extending to the weak links between decision and action (March 1991). The decision is then part of a trajectory where the direction is constantly adjusted by actors through the sensemaking process. According to this theory, sensemaking is initiated by "trigger events" and is constructed from preconceptions and the act of "speaking" (Weick et al. 2005). From this perspective, "organization" can be defined as "an attempt to order the intrinsic flux of human action, to channel it towards certain ends, to give it a particular shape, through generalizing and institutionalizing particular meanings and rules" (Tsoukas and Chia 2002, p. 570). Moreover, ambiguity—understood as the existence of different meanings that can be given to the available information (van Stralen 2015)—should be grasped through a mindful reworking of sense, in other words "to impose a plausible next step, but then to treat plausibility as both transient and as something compounded of knowledge and ignorance" (Weick 2015). At the extreme, the collapse of an organization, when workers run for their lives in response to a "cosmology episode," is explained by the sudden loss of sense and the impossibility of rebuilding it, following the collapse of the role structure (Weick 1993).

We argue in this chapter that none of these views of the concept of the decision or the action is sufficient to satisfactorily model how agents at the Dai Ichi plant acted. Thus, when asked, "What information led to your decision to return to the field?" (following the explosion of reactor 3) by members of the Commission appointed by the Japanese government, the superintendent of the plant, Masao Yoshida, replied:

> Rather than any particular information the fact was that, as far as reactor 1 was concerned, the explosion must have caused damage and there was a high risk that the water injection systems [were no longer functioning]. The immediate need was to check that no-one was dead and evacuate the wounded [...]. Once we had done this,

> we had to react. But everyone was in shock, frozen, unable to think. So I got them all together to talk to them. I told them how much I regretted having sent them back into the field when the danger had not passed, I had made the wrong decision and I asked them to forgive me. I also told them that it was very likely the water injection had stopped, […] if we did not react, we would have an even more catastrophic situation. But it was obvious that the ground was covered with rubble. First we had to clear this highly radioactive rubble. I told them to be careful […]. I asked them to do it, humbly. It was at that moment that I experienced one of the greatest emotions of my life. They all wanted to go back to the field, they even pushed each other out of the way.
>
> (Guarnieri *et al.* 2015, p. 259—our translation)

Understanding the decision process at the Fukushima site in the first week of the event requires comprehending the scope of such collective actions and the personal and interpersonal journey that led to them. We begin with a reminder of the accident scenario, then highlight the inadequacy of the conclusions of official bodies, and current theories that attempt to understand it. We then introduce the concept of the "extreme situation," as a theoretical framework for interpreting how the action unfolded.

Fukushima Dai Ichi: a reminder of the facts

On 11 March 2011, a strong earthquake strikes the plant, followed about 50 minutes later by a tsunami. All electrical generators are destroyed. Only DC power from reactor 3 and reactor 6's generator, used to cool reactors 5 and 6, are spared (Institut de radioprotection et de surêté nucléaire (IRSN) 2012; National Research Center (NRC) 2014). All of the lights go off, and the instrumentation is unusable. Vehicles and infrastructure are damaged, while debris blocks roads (National Diet of Japan Fukushima Nuclear Accident Independent Investigation Commission (NAIIC) 2012). It is difficult for the crisis unit, control rooms, and teams in the field to exchange information (IRSN 2012; NRC 2014). The crisis room is connected by videoconference to Tokyo Electric Power Company (TEPCO) headquarters in Tokyo, the main link with the outside.

Over the following days, operators try to cool the reactors and restore instrumentation used for monitoring. Yoshida orders an assessment be made of the different options for injecting water into the reactors, in anticipation of the failure of emergency cooling systems (Guarnieri et al. 2015). At the same time, control room operators venture out to inspect buildings. The superintendent quickly decides to use fire trucks and the fire circuit to supply fresh, and then seawater (opportunistically connected to available sources)—a maneuver not included in any emergency procedures—and operators use car batteries to power the instrumentation. However, they are only able to take readings intermittently, and must also rely on various other sources of information such as vibrations, noises, or smoke to infer the condition of machinery. Work is carried out in darkness, assessments of the state of buildings means learning their layout by heart, the heat quickly becomes unbearable, and aftershocks continue for several days.

Initially reactor 2 monopolizes everyone's attention as there is no information available to determine its condition. However, on the evening of 11 March, the level of radioactivity in reactor 1 increases. Yoshida understands that the water-level indicator, which, when it could be read, indicated a stable level, was stuck at a false value; that the emergency cooling system was actually ineffective; and that the core has not been cooled for several hours. Around midnight, the pressure in reactor vessel 1 is excessive; the decision is taken to manually perform the venting procedure in a highly radioactive environment. Water injection begins on 12 March at around 05:46, while venting is carried out at 14:30 (NRC 2014). At 15:36, the building of reactor 1 explodes, probably due to a hydrogen leak from the tank to the containment vessel into damaged pipes. Operators decide to return to the site to restore the water injection circuits destroyed by the explosion. At the same time, personnel in the crisis room hold discussions with headquarters about technical solutions to avoid the accumulation of hydrogen in the remaining buildings. The decision is taken not to intervene, given the extreme conditions at the site.

Hampered by both the difficulty of operations and the lack of staff, Yoshida gives the order to prepare for the venting of reactors 2 and 3, once reactor 1's injection circuit is finally ready. His overriding priority is to bring the situation in reactor 1 under control (Guarnieri et al. 2015). During the night, the state of reactor 3 can no longer be monitored; multiple efforts to depressurize the tank fail. On 13 March, it finally becomes possible to inject water via a fire protection system (NRC 2014), but reserves are exhausted at 12:20. They switch to seawater (NRC 2014). On 14 March at 11:01, a hydrogen explosion rocks reactor 3's building. Yoshida assembles his collaborators, who collectively decide to resume engineering work designed to restore the makeshift water injection circuits.

At dawn on 15 March, a strong earthquake is felt, accompanied by a loud noise and the fall of the pressure indicator in the tank of reactor 2. The origin of the explosion cannot be clearly established by agents who are on site. Although they give little credibility to the reading from the pressure indicator and, on a scientific level, the hypothesis that reactor 2 has exploded is inconsistent with the available information, Yoshida considers the noise to be the most important factor and orders an evacuation. Together with a few dozen employees, he remains to try to continue cooling operations (NRC 2014), people the media will call the "Fukushima 50" (Winerman 2011).

Recovery of control at the site required extensive engineering work, notably the delivery of seawater (NAIIC 2012). Due to the fear of contamination, logistical support and staff reinforcements remained particularly limited in the first few days. Despite these difficulties, the expectations of the Japanese population were immense and exacerbated by media coverage of the event (Cleveland 2014). The occupants of the plant were therefore directly exposed to reprimands from authorities (NAIIC 2012).

The decision under the microscope of investigators

While recognizing the immeasurable magnitude of the task of the agents on site, the reports of the Commission of Inquiry of the Japanese Diet (NAIIC 2012) and the commission established by the government (Investigation Committee on the Accident at the Fukushima Nuclear Power Stations of Tokyo Electric Power (ICANPS) 2012) conclude that the management of the accident was defective due to an erroneous assessment of the status of backup systems and inappropriate decisions, particularly with respect to priorities. They list a series of "diversions" from "rational" behavior, seen as the actions of professionals who progress in uncertain conditions on the basis of clear foundations.

For instance, one report states that

> faced with such an unimaginable situation and confused in the midst of a flood of information on the status of Units 1 to 6, however, neither the [on-site crisis cell] nor

> the [off-site crisis cell] thought to deduce the operational status of the IC [Isolation Condenser] from information related to the falling water level of Unit 1.
>
> (ICANPS 2011, p. 116)

Similarly, the Commission of the National Diet of Japan concluded that operators left the main control room to inspect the emergency cooling system of reactor 1 more than one and a half hours after its operational status became uncertain, and the main objective of the inspection was not even to confirm this status. Moreover, according to the investigator's assessment, operators quickly gave up inspecting the water level in reactor 1 because of a "small increase in the contamination level in the reactor building, despite the importance of the inspection" (NAIIC 2012, p. 26). The investigators also considered that due attention should have been paid to the depletion of batteries, in order to anticipate alternative water injection needs much earlier (ICANPS 2011).

Investigators were also surprised that operators had not considered the possibility of a hydrogen leak from the tank of the containment vessel into damaged pipes, although it was known that core fusion can produce large amounts of hydrogen. Yoshida answered that he had been aware that, if the core was damaged, hydrogen was produced; notwithstanding this, he felt that it would remain confined within the vessel, and he had focused on the threat of a container explosion, given the high pressure that had been observed. He stated,

> the top of the reactor building is covered and ventilation panels are arranged on the side. We had not even imagined that these panels were closed and that hydrogen and oxygen had accumulated. We focused on the containment vessel. […] We were prisoners of our *a priori* assumptions.
>
> (Guarnieri et al. 2015, p. 145—our translation)

Moreover, the entire international nuclear community was unaware of this scenario (ICANPS 2011).

Communication and the lack of a shared representation of the situation were also subject to the criticism that they led to dangerous decisions. For instance, aided by the partial restoration of the status display on the control panel, shift operators began to question the normal operation of the emergency cooling system of reactor 1 and eventually switched it off; however, this decision was not properly reported to the emergency response center for validation and coordination of further actions (ICANPS 2011).

These accident analyses create a form of contradiction. On the one hand, the extent of the damage severely constrained the ability of agents to act, particularly at the emotional level—something the investigators note when they state that "fear, stemming from a lack of information, caused mental stress among the workers and made the emergency response even more difficult" (NAIIC 2012, p. 17). On the other hand, the efficient conduct of operations would have required perfect coordination and a clear awareness of the situation and its ambiguities. Faced with this paradox, the conclusions that are drawn from these observations suggest a combination of "control" factors, i.e., systemic elements that are presumed to have a decisive influence on the management of a particular situation. Thus, according to the investigation commissions,

> individual operators are not to be blamed for these weaknesses. The underlying problems lie in TEPCO's organizational issues, such as the lack of nuclear safety

> preparedness against severe accidents—as exemplified by a lack of planning and implementation of adequate operator training, the lack of experience.
>
> (NAIIC 2012, p. 26)

From this perspective, future actions must in addition ensure the redundancy of measurement systems, the resistance of material resources, and their availability in the case of multiple failures.

Consequently, the lessons that have been taken from Fukushima Dai Ichi have focused on strengthening safety standards, while confirming the concept of "defense in depth" (NEA 2013, p. 24). The body of standards therefore provides both the scope of the risks and outlines the resources that must be implemented to manage and assess them in probabilistic terms. But this approach, in itself, does not foster an aptitude for "resilience" in complex socio-technical systems, which will continue to be affected by catastrophic events.[2] Paradoxically, deterministic approaches to safety dramatize uncertainty: when the aim is to create order, the introduction of disorder is destabilizing.

The system whose survival is threatened finds itself faced with a dilemma: it must find effective solutions despite the fact that its resources have been partially destroyed by the accident. In particular, engineering work can be seriously hampered by a lack of resources in an emergency that affects a whole society (Guarnieri and Travadel 2014). The Fukushima accident shows that the scenario ultimately depends on the ingenuity of people on site who are forced to adapt to unforeseen circumstances. Given this observation, the American National Research Council recommends improved modeling of the behavior and roles of agents in a serious nuclear accident (NRC 2014). However, current accident models only provide a fragmentary understanding of the mechanisms involved. By nature, they orient the analysis toward a discontinuity of production processes, the search for their causes and protective measures to be taken to mitigate their harmful effects. In turn, the agent's diagnosis, the corresponding procedure, and its effects are expected to fit perfectly into this causal sequence. *A contrario*, the difficulties faced by subjects who lack an applicable procedure and who must improvise solutions in a particularly hostile environment can be interpreted in terms of the processes that are inherent in the development of any collective activity and are highlighted in a crisis. In response to criticism from investigators regarding the confidence he placed in reactor 1's water-level indicators, Yoshida stated that

> anyway, in terms of solutions, we did not have anything much better than the diesel pump, injecting with the fire pump and using the fire engine, which we finally did. Could we have reacted more quickly if we had known? I think that physically, we could not have gone faster.
>
> (Guarnieri et al. 2015, p. 133—our translation)

Through this remark, Yoshida integrates his decision into a wider context of action and its impediments.

Modeling the decision and judging the rationality of an act are thus determined by the concepts of the agent and their "situation," "all of the concrete relations which, at a given moment, unite a subject or an associated group and the circumstances in which they must act."[3]

Engineering of the decision and collective action: the academic perspective

From a naturalistic point of view, the decision model is based on four postulates: (1) the decision maker is aware of alternatives; (2) they can assess the consequences of each of these options; (3) a set of values exists to guide them in the ordering of their preferences; and (4) there is a formal decision rule (March 1991). The process can be formalized via the optimization of a function (the utility) whereby a set of axioms express postulates about the properties of preferences (for a discussion of the various theories see Fishburn (1970). If it is not possible to make an objective assessment of the probabilities of the consequences of an act, the decision maker can use "subjective" probabilities. However, this assessment can be arbitrary especially as, in many practical situations, the behavior of agents does not reflect preferences that are consistent with these axioms (although this is not necessarily called an "error") (Gilboa and Marinacci 2013). To account for the assumed ignorance of the decision maker of certain states and their aversion to uncertainty, Gilboa and Schmeidler (1995) show that decisions can be seen as the maximization of a form of expected utility that takes into account the worst-case scenario. One of the core issues with this class of model is that judgments must be compatible with the entire web of beliefs held by the individual, even though "there can be no simple formal procedure for assessing the compatibility of a set of probability judgements with the judge's total system of beliefs" (Tversky and Kahneman 1974, p. 185).

Moreover, decision makers are proof of the limits of processing the available information and are often forced to rely on fragmentary data. The model is then that of a "bounded rationality" (March and Simon 1993, p. 190). In this approach, stress and emotions are seen as factors that influence the decision (for a review, see Kowalski-Trakofler et al. 2003), for example in the case of a threat to physical integrity and increased responsibility in the case of failure (Driskell and Salas 1991). However, these studies consider the stressor as an "ambient factor" that is independent of the performance of the action and external to the situation, while in an industrial crisis the decision is directly related to acting on the threatening process.

The analytical framework for investigators of the Fukushima accident is largely based on this economic view of the decision and the effect of stressors. The investigation becomes a search for deviations from axioms that are characteristic of safe behavior, or the acknowledgment of cognitive limitations. Probabilistic assessment is invoked as a way to judge the relevance of a decision, even though this judgment may be based on a subjective and retrospective examination of the facts. This criticism highlights a limitation that is more serious than it might initially seem. As Mugur-Schächter (2013, p. 3) argues, quoting Kolmogorov, "the mathematical representation of probabilities is not, contrary to [what the father of modern probability theory] previously believed, an abstract reformulation of a well-constructed factual concept of probability, but merely an interesting mathematical construct." For instance, there is no a priori justification for the mere existence of a factual probability in support of a probability law. "The probability laws used in scientific research, and even in engineering, are simply assumed to be involved as a consequence of the deliberately created stabilities, and are never actually known in advance" (Ibid., p. 58).

Observations of decision makers (notably firefighters) have made it possible to identify the processes involved in heuristic decisions (Natural Decision Making) when conditions change rapidly, tasks are not clearly defined, time is short and there are significant personal risks if there is an error (Klein and Klinger 2000). In these conditions, the decision maker's accounts of their decision making

> do not fit into a decision-tree framework; they are not "making choices", "considering alternatives", or "assessing probabilities", but they see themselves as acting and reacting on the basis of prior experience, generating and modifying plans to meet the needs of the situation.
>
> (Klein 1993, p. 139)

The decision maker acts on the basis of heuristics, then develops a mental simulation to assess the feasibility of the proposed response. This work is consistent with that of Gilboa and Schmeidler (1995), who axiomatized a Case-Based Decision Theory. However, these examples suggest that the strategy the decision maker adopts is the result of their experience (Lebraty et al. 2011) and can be taught (Volker 2011), in order to prepare them for a situation that is clearly "uncertain" but for which they are relatively equipped and which constitutes their predictable working environment.

Alternatively, the "environment" of the decision maker can be seen as a subjective construct. Schütz (1944) argues that the individual operates economically: the proposed action must simply be consistent with the typical elements of situations that have hitherto ensured that it is possible to carry out similar actions. The "crisis" is a weakness in the subject's coherent system, which ends when they can rebuild a set of reliable patterns of interaction that empower them to once again act socially. Thus, the actions of the decision maker can be seen as defining the framework for action through a formulation of the "situation" (Journé and Raulet-Croset 2008). In the case of Fukushima, such a model would negate the determining power of the physical conditions of action.

Ultimately, isolating the concept of the "decision" may seem artificial.[4] According to Weick (1993), contextual rationality is sensitive to the fact that social actors need to create and maintain inter-subjectively binding normative structures that sustain and enrich their relationships. He thus proposes to bypass the concept of decision and to move to the analysis of meaning through the sensemaking process at stake within organizations, "a process, prompted by violated expectations, that involves attending to and bracketing cues in the environment, creating inter-subjective meaning through cycles of interpretation and action, and thereby enacting a more ordered environment from which further cues can be drawn" (Maitlis and Christianson 2014, p. 67). The crisis is defined by the sudden occurrence of an improbable event that threatens this process (Weick 1993). This approach has been used to describe coordination activities given misleading information, an emergency context, or situations where the human (saving the patient in the case of emergency care practitioners) or social (saving one's reputation) stakes are high (Faraj and Xiao 2006).

Overall, the theory of the decision, including anchoring it in the sensemaking process, pays insufficient attention to the kinesthetic and emotional determinants of representations. An illustration of such a mechanism is found in Yoshida's testimony. He states that following the explosion of reactor 1 (when the pressure was about 500 kPa), the number "500" left him "ill at ease." He went on to add, "I know this is totally irrational, it was just a feeling" (Guarnieri et al. 2015, p. 256—our translation). More generally, the belief and value systems of both the agent and the collective are consistent with their immediate experience, their intuitive experience of what can actually be felt and confirmed by inter-subjectivity.

The "grip on reality" that authorizes action relates to a construction in which the subject's identity is forged. Asked about the legitimacy of his decision to inject seawater into the fire circuit, Yoshida replied, "for the first time, we were completely beyond any procedures [...]. So I thought I had to trust my instincts. [...] All these orders saying stop, do this, do that, they were all parasites" (Guarnieri et al. 2015, p. 192—our translation). He therefore anchors his actions as manager in his deepest self and highlights the boundary between a group in action and a surrounding instituted reality that the group emerges from.

The chasm between the objective-probabilistic assessments put forward by investigators and Yoshida's testimony, which is filled with emotions and perceptions can then be interpreted in the light of the work of Husserl (1976), and his observation that the life of the conscience is a life that completes a sense of being. According to the philosopher, the widespread application of this principle, including to science, undermines an approach of assigning to the world a mathematical-rational self, described by more-or-less complete theories.

In order to have a better understanding of the vital conflicts that the Fukushima Dai Ichi workers experienced when they made numerous unsuccessful attempts in responses to threats of social, mental, or even physical annihilation, we introduce the concept of the "extreme situation."

The concept of the extreme situation

We are principally interested in large-scale at-risk industries, in particular the nuclear sector, and specifically the management situation faced by operators in such an industry when they lose control of the facility. The "management situation" is understood here in the sense of Girin (1990); participants assemble and must accomplish, within a specified time frame, collective action leading to a result that is subject to external assessment. In an "extreme situation," the operator is regarded by society as being responsible for an intolerable risk with irreversible consequences, and agents must recover control of the installation in spite of any hazards.[5] Action therefore has to be interpreted through a complex of values that form the link between personal and social expectations.

At the most practical level, action consists of a set of operations or a series of gestures that are deployed in specific conditions. The resistance of reality initially manifests at this level and not only leads the subject to explore their options, but also to mislead themselves about their own ethical requirements. Caught between the physical impossibility of acting and incompatible requirements (orders to act, lack of procedures, etc.), the worker is torn by conflicting values. They must find a new motivation for their actions, i.e. they must alter their subjective investment in reality. This leads to a change in the meaning of action—action is here seen as a subordinate process to the representation of a result to be achieved (Leontiev 1977). Consequently, the reality of the activity becomes "what is not done, what is not done any more, but also what one attempts without success—the drama of failure—what one would have liked to do, what one thinks they would have been able to do otherwise" and "what one does to avoid doing what must be done" (Clot 2011, p. 17—our translation).

Unfolding action thus has a close link with the identity of subjects. Sigaut (2004) proposes a triangular relationship: *ego–reality–others*, upon which all action is based and identity is constructed on a daily basis. The shared experience of opposition to reality—and not the sharing of this experience—weaves the most basic social bonds, and reality only becomes powerful when it is socialized. The individual experiences reality by "doing," through the implementation of a social "tradition" (the *ego–others* link), which consists of a set of action models and values that the group uses to help or prevent each person building an effective relationship with reality. The identity thus requires authentication. It is indexed to reality and "the recognition of the identity by others, is always also the recognition by others of an authentic link between ego and reality" (Dejours et al. 1994, p. 3—our translation). The recognition is that of the quality of the result, acknowledged through the instrumented act. The contribution of the subject to the collective effort is demonstrated and becomes visible through specific gestures, which have a particular meaning for the collective and are evidence that their commitment is equal to that of others. "Without an adequate stage for their acts, without the material and social conditions for their implementation, actions immediately become unintelligible to others, or sources of misunderstanding" (Ibid., p. 5—our translation).

A more extensive interpretation of the link between individuals' internal conflicts and group action requires positing a media that connects individual desires and collective needs. In this respect, we draw upon Giust-Desprairies's (2009) theory of the "collective imaginary." The group is seen as an empirical unity of action centered on the realization of objectives as a function of resources. However, neither these objectives, nor the objects they relate to, can be reduced to a conscious representation. They are charged with affect and expectations, which gives rise to an imaginary scenario by the actors of their relationship to the object. The meeting between an instituted social universe that carries meaning and an individual who holds certain values leads to the emergence from the group of a specific imaginary content. An unstable equilibrium that is subject to incessant adjustments, this "collective imaginary" offers a grip on reality, a consensus on how to address reality that articulates a reduction between individual desires and collective achievement. The construction of the collective imaginary gives sense to the group identity, "which will be built on a sense of belonging emerging from a sharing of values, representations, beliefs from which the subjects will commit acts involving solidarity and cooperation to work toward the transformations they want from this reality" (Giust-Desprairies 2009, p. 131—our translation). Thus, through their actions, the subject feels as if they are participating in a story that is not their own. The activity is the production of objects and human relationships; it creates an ordered world in the realm of the imaginary.

We should note that this social mechanism is applicable to working situations, within a given institutional context. Even though, in general, it is not specific to cultural groups, the images formed within a particular collective imaginary are culture-driven. Indeed, the core of the group's imagination resonates with the encompassing social imaginary. These social representations pre-exist and can be subjectively accessed by individuals; therefore they govern the group's imaginary constructs.

In addition, the subject seeks acknowledgment of their identity from both their work colleagues and, at least as much, from their social environment. If these two entities have a different understanding of meaning, a conflict can emerge. By investing objects with particular values, the external view of the activity changes the economics of the situation. This judgment notably relates to the violation of certain values perpetrated by ego—and the alter-ego. The reality in question is therefore divided into imaginary objects where needs and fears are expressed. This disjuncture within reality releases vital tensions. As a corollary, "social alienation" is the fate of those whose reality is incomprehensible or unacceptable for others (Sigaut, 2004, p. 121—our translation). "Cultural alienation" is denial by the group of an uncomfortable reality—which is however, its *raison d'être*—motivated by an overriding imperative of solidarity among its members (Ibid.).

Subjects are therefore a construct that is in a perpetual state of suspension, which is forged and renewed through contact with the various elements of the situation. Artifacts occupy center stage in the social field; they create the theater for interactions and the generalization that is inherent in the establishment of the institution (Latour 2007). They determine the relationships that subjects have with themselves and their own identity construct, which the act emerges from. They weave an emotional connection between the individual and their employer (Rafaeli and Vilnai-Yavetz 2004). More generally, according to Leontiev (1977), mental representations of the objective world are governed by a process in which the subject enters into physical contact with it, and thereby obeys its intrinsic properties, its own relations. Recent studies show that the organism's bodily interaction with the environment is crucially important to its cognitive processes (Kerkhofs and Haselager 2006), and individuals produce concepts according to their perceptual experience (Niedenthal et al. 2009).

> It is impossible for a man not to be permanently changed and transformed by the sensory flow that runs through him. The world is the product of a body which translates it into perceptions and meaning, one cannot exist without the other. The body is a semantic filter.
>
> (Le Breton 2013, p. 47—our translation)

Moreover, from an anthropological perspective, the body can be seen as the primary technical object (Mauss, 1936). To "work," the subject mobilizes their subjectivity through their body, which

> we live in, which is experienced emotionally, the body that is also employed in the relation to the other: gestures, mimicry, […] the keys of a repertoire of bodily techniques […] dedicated to the expression of meaning and the desire to act on the awareness of the other.
>
> (Dejours 2001, p. 10—our translation)

The act therefore manifests according to three dimensions that define the "situation": the subject's subjective investment in reality, the collective construction of a world of representations in which work is judged, and society's opinion of the activity through the "instituted reality" in which the activity takes place. Sensory experience plays a special role as mediator in this process, in the development of concepts that are both individual and collective.

In an extreme situation, the sudden break with reality sweeps aside the group's collective imaginary. Decision processes become irrelevant, the action program must be reinvented and anchored in a new meaning into which subjects can invest themselves. Organizational theory tends to see the crisis as a situation "not only where resources are exhausted, but a situation that no-one has ever considered how to handle" (de Terssac 2011, p. 101—our translation). According to our theoretical framework, we can say that, in order to act, workers have *hic et nunc* to restore an enabling imaginary. However, when the production tool becomes unmanageable it not only creates serious constraints for workers but also, and this is one of the key elements of an extreme situation, for others who are not involved in production. It may lead simultaneously to the destruction of resources for an operator, a hostile physical environment, and material distress for the victims of a catastrophe. The worker is then faced with both the resistance of reality and potentially their social, mental, and sometimes physical annihilation. Each operation becomes an issue of survival or even a trade-off between one's own survival and that of others. The emergency emerges from how the work is viewed by the external body of society. Actions are therefore far-reaching, while at the same time the operator is without any effective means to take action, due to a lack of resources.[6]

The decision maker in an extreme situation

At Fukushima Dai Ichi, the unfolding events led to the simultaneous heating of three reactors and the explosion of two buildings. Operators were faced with a combination of overwhelming forces; the scene resembled a theater of war. In a situation that was beyond belief, action was focused on the recovery of the electricity supply, the verification (by any means possible) of the state of the reactors, and reducing pressure in the containment vessel. Achieving these goals required a series of complex operations (including venting) that were both far-reaching (e.g. implementing cooling systems through the transportation of sea water) and ingenious (e.g. the use of car batteries and the choice of water injection systems).

Action was made even more difficult by the lack of resources. In particular, firefighters working for TEPCO's subsidiary initially refused to take part in operations because of their fears for their health. The discord between the attitude of the plant's staff and that of firefighters is probably due to the tangential position of the latter with respect to the extreme situation, while social pressure to limit pollution was primarily applied to the operator. The redefinition of the action was therefore determined in part by the "proximal resources" that were available. This term refers particularly to artifacts that have associative links with others, where the usage is different but which have certain comparable or complementary physical properties, viewed from another angle. An illustration of this is the use of trucks equipped with concrete pumps. They were used to inject seawater directly from the roof of the damaged Fukushima Dai Ichi reactors; this was a world first that made it possible to contain fuel rod heating, and allowed time to install more permanent cooling systems (TEPCO, 2012).

The action program and decision making were shaped by the contact of subjects with this material and social reality, the physical challenges, their behavior, and mutual concern. Plunged into an unprecedented sensory universe, the operators at Fukushima Dai Ichi had to redefine meanings and values in their world. Actors opportunistically integrated "traditions" that were largely irrelevant to representations drawn from their immediate experience into this construction. Consequently, the available instrumentation was used as an anchor point to remodel representations. For example, the improvisation of an acceptable level of exposure to radioactivity before abandoning a mission was "materialized" through the "kamikaze" setting of dosimeters at MSv 100, with the result that the measurements that were taken no longer had any "objective" value, but became profoundly significant with respect to the level of commitment. Similarly, Yoshida expressed his "joy" when instruments indicated a "normal" water level in the tanks; however the instruments had already been proven to be erroneous several times before. This latter example should be interpreted less as an error on a scientific level, than as a fantastical representation of the situation that ultimately allowed a trade-off to be made between priorities for managing the reactors, at a time when human resources were dramatically lacking.

The mobilization of agents in this context did not happen just by itself. The concerns of staff only increased along with their fatigue. In this regard, supervisors reminded their subordinates of their responsibilities to the population, in particular their own families. Strengthening the feeling of group belonging through the idealization of action created the necessary foundation for the significant commitment of workers, similar to soldiers who fight as a "unit." The common ground shared by team members facilitated this development.

The behavior of leaders at all levels (the superintendent of the plant, control room supervisors, technical team leaders) appears to have been decisive in ensuring that action continued. Our understanding of the extreme situation encourages an exploration of the subject behind the manager, i.e. to recognize the unconscious, involuntary, non-manipulative aspects of their actions, and to admit that they are subject to social mechanisms, the influence of collaborators, and public opinion. The role of the manager can be thought of "as what is added to the prescription of power, management and control of the organization, the environment and oneself and which finds an arrangement with what is 'real' and that is different to it in every way" (Wolf-Ridgway 2010, p. 243—our translation). Their task is to reduce the lack of predictability; the extreme situation triggers a feeling of "uncertainty." Their decisions can of course be shaped by known biases in cognitive sciences (Tversky and Kahneman 1974). Nevertheless, above all, their approach is choreographed, dictated notably by defensive strategies, such as taking the stage to display a level of control that they do not have. For instance, the establishment of intervention teams at Fukushima is more consistent with a drama than the application of established procedures. Supervisors felt guilty when they had to designate personnel to check the reactor buildings (Kadota et al. 2014). Faced with this dilemma, some of them volunteered in the hope of triggering a collective movement, which is in fact what happened.

As workers intervened in increasingly severe conditions, group cohesion increased. They acknowledged each other's professionalism, and through this interaction they strengthened the community dimension. However, this idealized construction of an identity within the group on site contrasts sharply with the social disapproval experienced by these workers once they were outside. In fact, their motivation to act seems to have been anchored in the denial of this disapproval. Indeed, the construction of the collective imaginary is based on the eradication of the conflict and not its integration (Giust-Desprairies 2009). This resulted in an even more idealized action and its representation as a battle against the reactors, as indicated by Yoshida's frequent anthropomorphic references to them; he speaks of "three monsters," "three nuclear units that were unleashed," that must be "tamed" (Guarnieri et al. 2015, p. 202—our translation). The fantasy of action as a battle imputes an external cause to something workers would otherwise be responsible for. This tension probably explains the unusually high rates of psychological distress found among operators when they came face-to-face with the refugee population (Shigemura et al. 2012).

More generally, social pressure weighed heavily on the changing representations of the group on site. It was felt directly in the crisis room, through the video conference system and televisions. In particular, the hurtful statements of the Japanese Prime Minister on the morning of 15 March at TEPCO headquarters, mistakenly relayed to personnel in the crisis room, plunged them into disarray. These words contributed to Yoshida's decision when the three reactors appeared to be out of control at the same time. The "Fukushima 50," faced with their powerlessness, chose to transcend their social role and sacrifice themselves. The mark of a hero is to act

> for the good of a group when they have lost all other means to gain recognition. This is why the individual who sacrifices himself needs this group to define his own

> identity for him to agree to give the gift of his life. But it is also because he has already transformed his identity through an extension to the whole group that he has the courage to do it. The specific characteristic of the risk that is assumed is related to the value to the collective of this act. The latter in turn will give their identity to those who accept this risk.
>
> (de Mijolla-Mellor 2013, p. 12—our translation)

Conclusion

Heidegger (1971, p. 164) notes that "despite all conquest of distances the nearness of things remain absent." We therefore advance a hypothesis: it is because the nearness of things is distanced that the catastrophe, as "the irruption of the possible in the impossible" (Dupuy 2004, p. 10—our translation), reminds the Being of itself. The Dai Ichi occupants struggled to the death with things, things that were created and placed by technology behind control panels and the concrete walls of a nuclear reactor. To understand their action, how it unfolded, its meaning, its inhibitions, in short in order to draw lessons that reflect the grandeur of their gesture, we must abandon mentalist conceptions of a disembodied intellect, and particularly that of a decision maker in search of instrumental rationality. The "conscience implies for the subject their own body and a real social world" (Clot 2011, p. 22—our translation). It was above all, through their bodies that the Fukushima Dai Ichi operators avoided an unprecedented disaster. Their activity was fraught with life-threatening conflicts, to the point where they had to mediate between the survival of some and the sacrifice of others. Given these arguments, we propose a new interpretative framework, action in extreme situations, which reveals the human depths of the management of the accident on 11 March 2011.

The scenario of an industrial disaster depends largely upon the initiatives taken by workers on site (NRC 2014). Their ability to undertake collective action is a key factor to transitioning into resilience. In this respect, it has been shown that the principle of "on-scene initiative" fosters team performance in response to a disaster (GAO 2006, p. 11); in addition, it seems that strong political support facilitates flexibility within organizations (Pandey and Moynihan 2006). These considerations can be expressed within our conceptual setting in order to enunciate one principle and two resulting rules applicable to the sound management of extreme situations:

- *Fundamental principle.* The extreme situation is characterized by upheavals of the real and social holders of the collective imaginary; therefore, workers have to go beyond the collapse of their framework, create another, and act again as a group. Due recognition of this pattern by off-site managers is essential, even though a procedural rationality may be reassuring.
- *Rule #1.* In order to foster the emergence of meaning within the group, which is one that enables a new collective imaginary, leaders should provide a fixed point around which action can develop. They should do so as soon as they acknowledge the crisis, and even if, in the end, it may turn out that alternative solutions would have been more optimal.
- *Rule #2.* As a corollary to Rule #1, on-site workers—including managers—should be given sufficient freedom to act, not only because they are in the best position to identify the opportunities for action and enact solutions, but also because this is one of the conditions for them to subjectively invest in their action and to commit themselves to self-threatening actions through participation in the group.

Notes

1 Source: *Dictionnaire culturel*, A. Rey, (Ed.) Le Robert, Paris, 2005, entry "decision" (our translation).
2 Generally speaking, the "resilience" of a system can be defined as its intrinsic ability to adjust its functioning prior to, during, or following changes or disturbances, so that it can sustained required operations under both expected and unexpected conditions (Hollnagel et al. 2006).
3 Source: *Dictionnaire culturel*, A. Rey, (Ed.) Le Robert, Paris, 2005, entry "situation" (our translation).
4 For instance, the "Garbage Can Model" (Cohen et al. 1972) of the decision, which is applicable when there is no clear preference, no method, and a variable number of participants, represents the organization as a vector for problems, which contains solutions and sometimes finds itself faced with many different opportunities, although these elements occur together randomly.
5 Some "management situations" as described by Girin, are referred to as "extreme" when they present three characteristics: time sensitivity, uncertainty, and risk (Lebraty et al. 2011). This definition is applied in particular to interventions in a clearly changing context—such as practitioners in an emergency care center or mountaineers—but that constitute the predictable working environment of the agent. It therefore does not capture the crucial conflicts that Fukushima Dai Ichi workers had to deal with.
6 In this context, resources refer to material and human resources, know-how, methods, and procedures or, if they are lacking, the time to acquire them.

References

Cleveland, K. (2014) "Significant Breaking Worse," *Critical Asian Studies*, 46(3), pp. 509–539.

Clot, Y. (2011) "Théorie en clinique de l'activité," in Maggi, B. (Ed.) *Interpréter l'agir*. Paris: PUF, pp. 17–39.

Cohen, M. D., March, J. G., and Olsen, J. P. (1972) "A Garbage Can Model of Organizational Choice," *Administrative Science Quarterly*, 17(1), pp. 1–25.

Dejours, C. (2001) "Subjectivité, travail et action," *La Pensée*, 328, pp. 7–19.

Dejours, C., Dessors, D., and Molnier, P. (1994) "Comprendre la résistance au changement," *Documents pour le médecin du travail*, 58, pp. 1–8.

Driskell, J. E. and Salas, E. (1991) "Group Decision Making Under Stress," *Journal of Applied Psychology*, 76(3), pp. 473–478.

Dupuy, J.-P. (2004) *Pour un catastrophisme éclairé*. Paris: Seuil (Points essais).

Faraj, S. and Xiao, Y. (2006) "Coordination in Fast-Response Organizations," *Management Science*, 52(8), pp. 1155–1169.

Fishburn, P. C. (1970) *Utility Theory for Decision Making*. New York: John Wiley & Sons.

GAO (2006) *Coast Guard. Observations on the Preparation, Response, and Recovery Missions Related to Hurricane Katrine*. Report to Congressional Committees GAO-06-903. Washington, DC: US Government Accountability Office.

Gilboa, I. and Marinacci, M. (2013) "Ambiguity and the Bayesian Paradigm," in Acemoglu, D., Arellano, M., and Dekel, E. (Eds.) *Advances in Economics and Econometrics. Tenth World Congress*. Cambridge, UK: Cambridge University Press (Econometric Society Monographs), pp. 179–242.

Gilboa, I. and Schmeidler, D. (1995) "Case-Based Decision Theory," *The Quarterly Journal of Economics*, 110(3), pp. 605–639.

Girin, J. (1990) "L'analyse empirique des situations de gestion: éléments de théorie et de méthode," in Martinet, A. C. (Ed.) *Epistémologies et sciences de gestion*. Paris: Economica, pp. 141–182.

Giust-Desprairies, F. (2009) *L'imaginaire collectif*. Toulouse: ERES (Poche-Société).

Guarnieri, F. and Travadel, S. (2014) "Engineering Thinking in Emergency Situations: A New Nuclear Safety Concept," *Bulletin of the Atomic Scientists*, 70(6), pp. 79–86.

Guarnieri, F., Travadel, S., Martin, C., Portelli, A., and Afrouss, A. (2015) *L'accident de Fukushima Dai Ichi: le récit du directeur de la centrale. Volume I, L'anéantissement*. Paris: Presses des Mines.

Heidegger, M. (1971) "The Thing," in *Poetry, Language, Thought*. Translated by A. Hofstader. New York: Harper and Row, pp. 163–180.

Hollnagel, E., Woods, D. D., and Leveson, N. (2006) *Resilience Engineering: Concepts and Precepts*. Aldershot, UK: Ashgate Publishing.

Husserl, E. (1976) *La crise des sciences européennes et la phénoménologie transcendentale*. Translated by G. Granel. Paris: Gallimard (Collection Tel).

Institut de radioprotection et de surêté nucléaire (IRSN) (2012) *Fukushima, un an après. Premières analyses de l'accident et de ses conséquences*. IRSN/DG/2012-001. Fontenay-aux-Roses: IRSN.

Investigation Committee on the Accident at the Fukushima Nuclear Power Stations of Tokyo Electric Power (ICANPS) (2011) *Interim Report*. Available at: http://www.cas.go.jp/jp/seisaku/icanps/eng/interim-report.html.

Investigation Committee on the Accident at the Fukushima Nuclear Power Stations of Tokyo Electric Power (ICANPS) (2012) *Executive Summary of the Final Report*. Available at: http://www.cas.go.jp/jp/seisaku/icanps/eng/final-report.html.

Journé, B. and Raulet-Croset, N. (2008) "Le concept de situation: contribution à l'analyse de l'activité managériale en contextes d'ambiguïté et d'incertitude," *M@n@gement*, 11(1), pp. 27–55.

Kadota, R., Tokuhiro, A., and Varnam, S. (2014) *On the Brink: The Inside Story of Fukushima Daiichi*. Fukoka, Japan: Kurodahan Press.

Kerkhofs, R. and Haselager, W. F. (2006) "The Embodiment of Meaning," *Manuscrito*, 29, pp. 753–764.

Klein, G. A. (1993) "A Recognition-Primed Decision (RPD) Model of Rapid Decision Making," in Klein, G. A., Orasanu, J., Calderwood, R., and Zsambok (Eds.) *Decision Making in Action: Models and Methods*. Norwood, NJ: Ablex Publishing Corporation, pp. 138–147.

Klein, G. and Klinger, K. (2000) "Natural Decision Making," *Human Systems IAC Gateway*, 11(3), pp. 16–19.

Kowalski-Trakofler, K. M., Vaught, C., and Scharf, T. (2003) "Judgment and Decision Making Under Stress: An Overview for Emergency Managers," *International Journal of Emergency Management*, 1(3), pp. 278–289.

Latour, B. (2007) "Une sociologie sans objet? Remarques sur l'interobjectivité," in Debary, O. and Turgeon, L. (Eds.) *Objets et mémoires*. Paris: Editions de la Maison des sciences de l'homme, pp. 37–57.

Lebraty, J.-F., Lièvre, P., Récopé, M., and Rix-Lièvre, G. (2011) "Stress and Decision. The Role of Experience," in *10th International Conference on Naturalistic Decision Making. 22 May–3rd June 2011*. Orlando, Florida: University of Central Florida, pp. 266–270.

Le Breton, D. (2013) *Anthropologie du corps et modernité*. Quadrige. Paris: Presses Universitaires de France (Sociologie d'aujourd'hui).

Leontiev, A. N. (1977) "Activity and Consciousness," in *Philosophy in the USSR, Problems of Dialectical Materialism*. Moscow: Progress Publishers, pp. 180–202.

Maitlis, S. and Christianson, M. (2014) "Sensemaking in Organizations: Taking Stock and Moving Forward," *The Academy of Management Annals*, 8(1), pp. 57–125.

March, J. G. (1991) "How Decisions Happen in Organizations," *Human-Computer Interaction*, 6, pp. 95–117.

March, J. G. and Simon, H. (1993) *Organizations*. 2nd edn. Cambridge, UK: Wiley-Blackwell.

Mauss, M. (1936) "Les techniques du corps," *Journal de psychologie*, 32(3–4), pp. 365–86.

de Mijolla-Mellor, S. (2013) "La mort du Héros," *Topique*, 125(4), pp. 7–17.

Mugur-Schächter, M. (2013) "On the Concept of Probability," *Mathematical Structures in Computer Science*, 24(Special Issue 3), pp. 1–91.

The National Diet of Japan Fukushima Nuclear Accident Independent Investigation Commission (NAIIC) (2012) *The Official Report of the Fukushima Nuclear Accident Independent Investigation Commission. Main Report*. Available at: http://warp.da.ndl.go.jp/info:ndljp/pid/3856371/naiic.go.jp/en/report/.

National Research Council (NRC) (2014) *Lessons Learned from the Fukushima Nuclear Accident for Improving Safety of U.S. Nuclear Plants*. Washington, DC: The National Academies Press.

Nuclear Energy Agency (NEA) (2013) *The Fukushima Daiichi Nuclear Power Plant Accident: OECD/NEA Nuclear Safety Response and Lessons Learnt*. Special Report 7161. Paris: OECD.

Niedenthal, P. M., Winkielman, P., Mondillon, L., and Vermeulen, N. (2009) "Embodiment of Emotion Concepts," *Journal of Personality and Social Psychology*, 96(6), pp. 1120–1136.

Pandey, S. K. and Moynihan, D. P. (2006) "Bureaucratic Red Tape and Organizational Performance: Testing the Moderating Role of Culture and Political Support," in Boyne, G. A., Meier, K. J., O'Toole Jr., L. J., and Walker, R. M. (Eds.) *Public Service Performance*. Cambridge, UK: Cambridge University Press, pp. 1–36.

Rafaeli, A. and Vilnai-Yavetz, I. (2004) "Emotion as a Connection of Physical Artifacts and Organizations," *Organization Science*, 15(6), pp. 671–686.

Rey, A. Ed. (2005) *Dictionnaire culturel*, Paris: Le Robert.

Schütz, A. (1944) "The Stranger: An Essay in Social Psychology," *American Journal of Sociology*, 49(6), pp. 499–507.

Shigemura, J., Tanigawa, T., Saito, I., and Nomura (2012) "Psychological Distress in Workers at the Fukushima Nuclear Power Plants," *Journal of American Medical Association*, 308(7), pp. 667–669.

Sigaut, F. (2004) "Folie, réel et technologie. À propos de Philippe Bernardet, Les Dossiers noirs de l'internement psychiatrique, Paris, Fayard, 1989," *Travailler*, 12(2), pp. 117–130.

van Stralen, D. (2015) "Ambiguity," *Journal of Contingencies and Crisis Management*, 23(2), pp. 47–53.

Tokyo Electric Power Company (TEPCO) (2012) *Accident Overview at Fukushima Daiichi NPS, Fukushima Daiichi—A One Year Review*. Available at: http://www.tepco.co.jp/en/nu/fukushima-np/review/review1_2-e.html.

de Terssac, G. (2011) "Théorie du travail d'organisation," in Maggi, B. (Ed.) *Interpréter l'agir. Un défi théorique*. Paris: PUF (Le travail humain), pp. 97–121.

Tsoukas, H. and Chia, R. (2002) "On Organizational Becoming: Rethinking Organizational Change," *Organization Science*, 13(5), pp. 567–582.

Tversky, A. and Kahneman, D. (1974) "Judgement under Uncertainty: Heuristics and Biases," *Science*. (New Series), 185(4157), pp. 1124–1131.

Volker, F. (2011) "Decision-making under Uncertainty: Using Case Studies for Teaching Strategy in Complex Environments," *Journal of Military and Strategic Studies*, 13(2), pp. 1–21.

Weick, K. E. (1993) "The Collapse of Sensemaking in Organizations: The Mann Gulch Disaster," *Administrative Science Quarterly*, 38(4), pp. 628–652.

Weick, K. E. (2015) "Ambiguity as Grasp: The Reworking of Sense," *Journal of Contingencies and Crisis Management*, pp. 1–7.

Weick, K. E., Sutcliffe, K. M., and Obstfeld, D. (2005) "Organizing and the Process of Sensemaking," *Organization Science*, 16(4), pp. 409–421.

Winerman, L. (2011) "The 'Fukushima 50': Nuclear Workers Stay Behind to Brave Plant's Woes," *PBS NewsHour*, 16 March. Available at: http://www.pbs.org/newshour/rundown/nuclear-plant-workers-become-anonymous-faces-of-bravery/.

Wolf-Ridgway, M. (2010) *Les apports de la clinique du travail à l'analyse de la "présentation de soi" chez le dirigeant d'entreprise*. Paris: Conservatoire national des arts et metiers (CNAM).

5 Prevention versus response

Application in the management of disease

Amy Hagerman, Bruce McCarl,
Aklesso Egbendewe-Mondzozo, and Levan Elbakidze

Introduction

Incidents of disease, like other kinds of extreme events, can be costly and are generally uncertain in incidence, location, and severity. Even defining an event as "extreme" is complex. An event could be called "extreme" when it occurs at all if it is a novel or low-probability event. Alternatively, it could be identified as "extreme" when high damages, numbers of infections, and/or response costs are involved. Nevertheless, given that an event is perceived to be extreme and is characterized by ambiguities and uncertainties, decision makers must choose where to operate on the continuum between taking

- ex-ante actions—spending resources on preventing it from occurring or investing in response capabilities or
- ex-post actions—waiting until it occurs before undertaking response and recovery activities.

To put the ex-ante versus ex-post action decision in context, consider the 2009 novel H1N1 outbreak in the United States. Many would argue this was an extreme event as there were approximately 43,677 laboratory-confirmed cases and 300 deaths from April 2009 to July 2009 (Shrestha et al., 2011). This event, however, was the subject of ex-ante action when the Centers for Disease Control (CDC) created a pandemic influenza response plan, a Strategic National Stockpile (SNS) of equipment and medications, and a training program consisting of disease response exercises based on a potential major H5N1 event. Also, ex-post in April 2009, the CDC began working on H1N1 vaccine development resulting in a vaccine that was deployed in October 2009 (CDC, 2010).

The above case illustrates a situation in which decisionmakers, both public and private, chose to address a class of extreme event by taking actions across the spectrum of prevention, preparedness, response, and recovery (PPRR) activities. Here we characterize the economics of such a decision using a "balance" model (Elbakidze & McCarl, 2006) that blends ex-ante prevention and preparedness activities (PP) investments with uncertain ex-post response and recovery (RR) expenditures with the ultimate goal of minimizing the total costs of a disease outbreak.

Ex-ante actions in addressing possible extreme disease events generally involve investments in events that may never occur; ex-post actions involve certainty about the occurrence but may be hampered by lack of ex-ante PP activities. Furthermore, the ex-ante activities, whether chosen or foregone, are generally irreversible at the time of an occurrence. The balance model framework accounts for this setting, modeling the sunk cost and irreversible nature of ex-ante actions and the uncertainty surrounding the occurrence of a disease event, as well as the magnitude of losses as a function of PP and RR activities. In this chapter we focus on motivation, basic theory, and prior applications of the balance model plus a specific example in the context of a high-consequence, low-probability animal disease outbreak of avian influenza in the United States. Although the human consequences of H5N1 will not be explicitly examined in the example, the framework could be used similarly to illustrate allocation of efforts across PP and RR activities in the context of zoonotic diseases, recognizing that estimating the damages of human illness is a separate, multifaceted issue (Lichtenberg, 2010).

Factors influencing investment decisions under uncertainty

Probabilistic risk assessment (PRA) has been advocated and used in many environmental health hazard contexts (Lichtenberg, 2010). Such assessments explicitly consider uncertainty using probability distributions of estimated risks. Monte Carlo methods are used to combine probability distributions of uncertain parameters into an analysis system to produce probability distributions of estimated risk as affected by risk mitigation strategies. One can set up a model with a set of ex-ante investments and responses then solve it to find the breakeven event probabilities that justify various degrees of ex-ante actions (Elbakidze & McCarl, 2006; Egbendewe-Mondzozo et al., 2013).

Investment decisions in PPRR activities are influenced by a variety of factors, including biophysical realities, economics, politics, behavioral considerations, social acceptability, and scarcity of resources. Economically, these factors alter the long-term damages when outbreaks occur plus the ex-ante PP costs. The ex-ante PP costs include many items such as direct costs associated with periodic surveillance and inspection, costs of stricter production standards and regulations, costs of vaccination including vaccine production and administration, costs of constructing labs, training response forces, etc. Post-event damages from an outbreak would include losses in animal sales, producers' income, and consumers' welfare due to supply reduction or consumption pattern change (Rich et al., 2005; Jin et al., 2009; Kinsey et al., 2011) plus trade effects (Paarlberg & Lee, 1998).

Investments in ex-ante preparedness and prevention activities are often sunk costs, unrecoverable at a later time, and any earnings those resources could have generated in alternative uses are foregone as opportunity costs of investment in PP. However, once resources have been dedicated to PP activities, the result is a cost-reducing effect on RR when outbreaks occur plus possible co-benefits in other settings. Such was the case in the H1N1 example, where influenza preparedness investments initially made with H5N1 in mind were used in H1N1 response. In the contexts of uncertain incidence, severity of outbreak occurrence expenditures on PP activities act as risk-reducing insurance actions.

Prevention activities focus on reducing or eliminating the probability of the adverse event. For many illnesses, prevention activities can range from vaccinations in at-risk areas to biosecurity practices as simple as regularly washing hands. Preparedness activities are those that either enable or enhance the effectiveness of response activities once an event occurs. Some examples of preparedness activities are surveillance and detection capability enhancement, responder training and education, building diagnostic capabilities, stockpiling vaccine, screening of imported animals, sanitation guidelines and protocols, and inspection and testing of biological materials (for example, semen). Prevention can also occur in other ways, such as through investment in disease eradication initiatives in the countries with endemic disease statuses. Such activities were crucial in the U.S. response to H1N1.

Ex-post event response activities are aimed at minimizing the damages of the event after the event takes place. Examples include infected animal slaughter and disposal, quarantines, and accelerated vaccination, movement bans, enhanced detection, and animal tracing. Recovery activities are efforts which seek to rebuild the system to its pre-event condition or even a better state. Examples of ex-post recovery activities are industry efforts to rebuild consumer confidence in food safety after the event, restocking subsidies, indemnity payments, post-event surveillance, and zoning of production regions for the purpose of reopening animal product exports. The effectiveness of the response and recovery activities depends on resilience of the system which in part depends on investments made in preparedness pre-event.

Animal disease background

Some animal diseases are characterized by high morbidity (sickness) and mortality rates along with the capacity to spread rapidly through animal populations irrespective of geographic borders and are called transboundary animal diseases or TADs.[1] Examples of TADs include foot-and-mouth disease (FMD), bovine spongiform encephalopathy (BSE), and highly pathogenic avian influenza (HPAI). Such diseases can cause significant economic and productivity losses, at least partially through trade embargoes. Some diseases such as H5N1can be transmitted to humans (zoonotic diseases), having human health consequences. These damages can stretch across many years and regions but the outbreak occurrence is typically rare and PP activities can be put in place to reduce the probability of occurrence and event severity. Due to this low probability of occurrence and uncertain but potentially large event costs, it is often difficult to decide how much investment in PP measures is justified. The challenges of incorporating both uncertain damages and uncertain probabilities of event occurrence are discussed later in the chapter.

In the post 9/11 era in the United States, a call for greater investment in PP for biosecurity was made to counteract threats from intentional as well as unintentional TAD events (Jin et al., 2009). However, limited resource availability for PP investment against the vast array of threats implies the need for a framework for prioritization of threats and investments in mitigation capacity.

To date, preventive measures in the United States have been largely successful in keeping out many economically damaging animal diseases. For example, FMD has not been detected in the United States since 1929. Also, efforts addressing one disease may also help avoid incidents of other diseases (for example, surveillance and testing protocols for one TAD might catch others).

Ongoing animal disease preparedness efforts in the United States include investments in the National Veterinary Stockpile, building laboratory capacity within the National Animal Health Laboratory Network (NAHLN), and the animal traceability system (ADT). Some PP activities are specific to a single disease of concern, such as vaccine stockpiles, whereas others apply to a variety of TADs, such as the NAHLN and ADT.

Basic theory: balancing pre-event costs of prevention and preparedness and post-event costs of response and recovery

In order to decide how to allocate effort between ex-ante and ex-post activities, decision makers need to form expectations about outbreak probabilities and about costs of the outbreak given ex-post RR activities as well as how the outbreak costs are affected by various levels of ex-ante PP investment and by various levels of RR activities. Expected values as well as variability of outbreak costs contingent on PP and RR activities play important roles in decision making.

Forming reasonable cost expectations is difficult due to the infrequency with which TAD outbreaks occur. TAD outbreaks, like many other types of extreme events, are highly uncertain in nature. For example, FMD has not been confirmed in the United States since 1929, so little information is available on which to base an outbreak probability (Elbakidze et al., 2009). Furthermore, the United States is closely connected with the rest of the world through extensive movements of people, animals, and products, and this poses numerous challenges in terms of assessing outbreak probabilities and managing TAD threats (Torrey & Yolken, 2005).

In most cases PP activities cannot eliminate all possibilities of an outbreak. Also, the potential size and duration of any TAD outbreak is highly uncertain. Both observed and simulated outbreaks have shown that disease spread can vary widely, more so for some diseases than others. For example, simulated FMD introductions in the United States show a wide range of total impacts in the billions of dollars depending on the

- introduction scenario;
- speed and effectiveness of response;
- duration and geographic extent of the outbreak;
- extent of embargoes on trade placed by trading partners; and
- the reaction of consumers to the presence of disease and to control measures.

U.S. simulations show moderate outbreaks of less than one calendar quarter duration with trade embargoes of less than one year on livestock and livestock products and little negative consumer reaction lead to losses of less than $5 billion in the livestock industry (Pendell et al., 2007; Paarlberg et al., 2008). However, outbreaks lasting longer than one quarter or with extended trade embargoes and marked negative consumer reaction can lead to economic loss estimates which exceed $100 billion (Hayes et al., 2011; Oladosu et al., 2013). Most studies that assume a major export embargo in the year an outbreak occurs fall somewhere in the middle.

Cost expectations are influenced by past experiences under similar types of events in the country of concern or in other countries and can also be informed by simulation models. For example, the costs of responding to the 2002–03 exotic Newcastle disease (END) outbreak that occurred in California were $161 million and involved the depopulation of 3.16 million birds (CDFA, 2014). Understanding the factors that contributed to the cost of response in this outbreak, particularly the high rate of incidence in backyard flocks, helps project the sort of costs that might be incurred from a response to END or HPAI elsewhere in the United States.

Another factor that affects the ex-ante/ex-post balance is cost sharing and response effort as it is distributed among affected parties. Centralizing effort and spending allows for cooperative risk-management strategies, where the goal is the reduction of total impacts across all affected groups. If the effort is more individual based then individual parties may choose to do very little in hopes of freeriding on PP investments made by other parties, which may lead to underinvestment in PP activities (Gramig et al., 2006, 2009). Gramig and others (2006, 2009) discussed the role indemnity compensation plays as an incentive for biosecurity investment and rapid reporting of disease outbreaks, or even a disincentive if improperly structured. A complementary way to examine the issue of preventing freeriding is to analyze PP investments via a balance model, where coordinated activities via government regulatory programs or incentive compatible mechanisms remove freeriding potential.

The balance model

The conceptual economic model of this ex-ante/ex-post effort balancing problem is the balance model of Elbakidze (2004) and Elbakidze and McCarl (2006). That model divides the time horizon of decision making into two stages (Figure 5.1). The first stage is the ex-ante investment opportunity stage. There, investments in PP activities, such as establishing minimum quality standards, screening of imported and/or marketed animals, conducting surveillance, enhancing detection capability, and stockpiling vaccines, can be made. These investments facilitate mitigation either by lowering outbreak probability, lowering damages, or increasing the effectiveness of RR activities relative to what would have occurred without the ex-ante investments in PP. Evaluating the return on PP investment is made difficult when PP investments are not disease specific, since one must decide how to attribute costs across various diseases.

The second stage is ex-post after an outbreak could have occurred and contains both a "no outbreak" state and cases where outbreaks occur of possibly different levels of intensity where RR activities are needed. Response measures may include actions such as evacuation, depopulation, movement controls, quarantine, and vaccination. Clearly, the effectiveness of many response activities depends on whether PP investments were made in the first stage. The more investments in response capability, disease prevention, or rapid detection that are made in the first stage the more effective are some of the RR activities in the second stage. For example, the effectiveness of vaccination as a response strategy depends on investments made in vaccine availability (Elbakidze et al., 2009; Egbendewe-Mondzozo et al., 2013). Similarly the ability to rapidly identify potentially infected premises and impose early movement bans or depopulation depends in part on investments made in traceability (Elbakidze, 2007).

A graphical portrayal of the logic underlying the balance model is represented in Figure 5.2 (Elbakidze, 2004). The graph shows how ex-ante, ex-post, and total costs react to

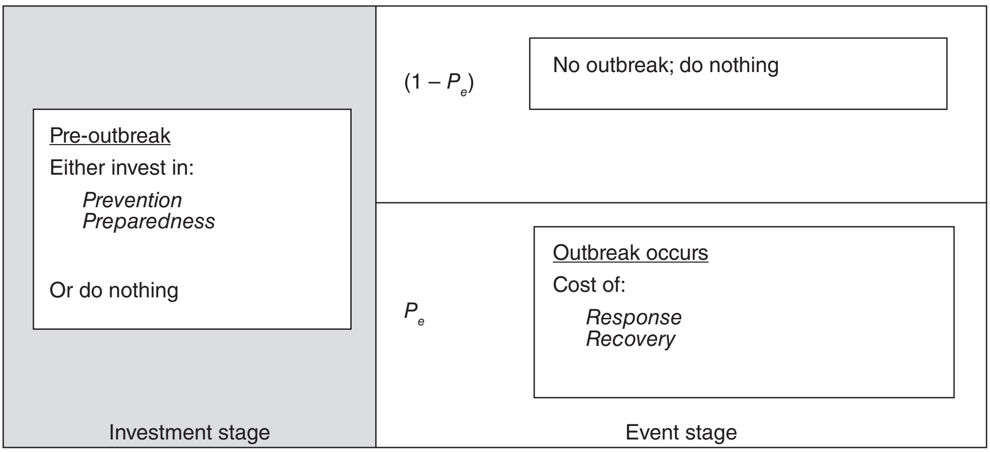


Figure 5.1 Event probability and decision-making stages.

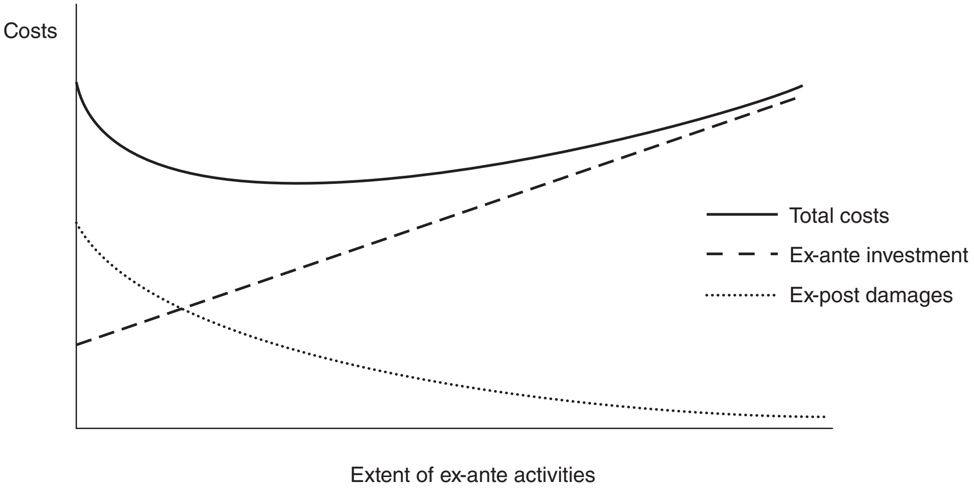


Figure 5.2 Conceptual balance model.

changes in the levels of ex-ante activity. The more activities are adopted ex-ante, the higher the ex-ante costs and the lower the ex-post costs. Generally, the ex-ante investments are irreversible in the sense that they are made prior to the outbreak and are incurred whether or not an outbreak ever happens. On the other hand, ex-post RR costs are weighted by the probability of the outbreak. In this economic approach the objective of the decision maker would be to minimize the expected total costs of the event by deciding how much to invest in ex-ante PP activities plus the nature of the RR activities to undertake given an outbreak occurs and the decision maker may be willing to tolerate higher expected total costs in return for reduced variability in potential total costs (Elbakidze et al., 2009; Hagerman et al., 2012).

Model set up

The balance model requires information on

- available options for ex-ante PP activities as well as the options for RR activities with associated costs and effectiveness measures;
- the expected level of ex-post RR and disease costs if those investments *were not* made; and
- the expected level of ex-post RR and disease costs if investments *were* made in PP activities.

Important parameters in the balance model pertain to the probability of event occurrence, severity of the event, and effectiveness of available ex-ante and ex-post mitigation activities. In TAD mitigation studies, these uncertainties are most often addressed using stochastic parameters for disease introduction, spread, and control effectiveness in an empirical epidemic model. However, uncertainties also persist in other parameters. For example, reaction of consumers to disease introduction and/or food contamination varies depending on contexts and RR activities. Perhaps the most problematic parameter is the probability of a TAD introduction. One way to apply the above balance model is to estimate the probability of event occurrence at which the ex-ante investments are justified, that is the probability where the ex-ante costs are just offset by the reduction in expected level of damages plus RR costs.

Mathematically a simple form of the balance model under risk aversion (in the framework of a mean variance portfolio selection model) can be represented as follows.

$$Min\ Cost(PP, RR) = FC * Y + P_S(Y) * [VCM_I * RR_I + VCM_{NI} * RR_{NI} + DV * N(Y, RR_I, RR_{NI})] + RiskAver * Var(Costs(Y, RR_I, RR_{NI}))$$

$$Subject\ to: -CAP * Y + RR_I \leq 0$$

$$RR_I = RR_{NI} = 1$$

The objective function is minimization of expected cost plus a risk-aversion parameter times the variability of costs. The formulation includes an irreversibility constraint for the sunk investment in PP. Let be a decision variable that takes a non-zero value if indeed PP investment activities are made. Let be the sunk cost incurred by the PP investment activity and CAP the capacity of those investments if made to support response and recovery, for example, the cost of building laboratory facilities for animal disease testing and the associated capacity, or the costs of stricter biosecurity and quality regulations and their capacity. These costs are incurred even without a disease event. In turn, let RR_I be the response and recovery activities employed if the PP investments were made and RR_{NI} be the response and recovery activities employed if the PP investments were not made. Also, here VCM_I denotes the variable costs of RR_I activities such that the cost of applying vaccines or using the diagnostic labs or operating with trained personnel and VCM_{NI} denotes the variable costs of RR_{NI} activities. $N(Y, RR_I, RR_{NI})$ is spread of the disease (say in terms of infected animals) as a function of response as affected by the preparedness activities. One can assume that $dN/dRR_I<0$, $dN/dRR_{NI}<0$, and $d^2N/dRR_I dY<0$, meaning that more response activities reduce the spread of the disease and the PP investment increases the effectiveness of the response activities. Finally, DV is the value of damages per unit of spread or per animal infected.

The objective function also includes a probability for the event, where P_ε (Y) is the probability of the event occurring as a function of prevention activities. Here, we assume when investment is made in PP activities it reduces the probability of event occurrence, $d(P)/d(Y)<0$.

The objective function also includes a component for risk preference of the decision maker. There the term $var(costs(Y,RR_I,RR_{NI}))$ represents the variance in response plus spread damages given the extent of ex-ante investment in PP and the response and recovery alternatives. The more risk averse the decision maker, the more weight is assigned to variability of potential total costs.

The first constraint imposes an irreversibility condition that allows the capacity of the PP investments to support the RR_I activities to be used only if the PP investments are made in the first stage. The second constraint mandates a response given the outbreak. There may be applications of the model where one may not wish to mandate any activity, meaning it may be more economically optimal to not undertake either ex-ante or ex-post activities when damages are minor and PP and RR is relatively expensive. However, for the purposes of extreme disease response evaluation, we felt this was a reasonable restriction to make.

The balance model illustrates the tradeoffs with regards to ex-ante investments in PP activities with reductions in damages under an outbreak along with the fixed and variable cost/irreversible nature of the PP investments and the uncertain nature of the ex-post outcomes. Overall, the model takes on the goal of minimizing total expected costs of implementing PP actions plus the probabilistically weighted damages given RR when the outbreak occurs plus weighted penalty for variability in expected costs.

Application of the balance model to avian influenza vaccination investment

One threat many countries face is the possibility of an HPAI outbreak. In the face of such threats, a number of countries have been investing in poultry vaccine stockpiles. Here we investigate the issue in a balance model.

Background on vaccines and HPAI

Vaccines have been stocked against the possibility of an HPAI outbreak. For example, in 2005 the United States contracted with Fort Dodge Animal Health, a vaccine development company, to produce 40 million doses of vaccine for several subtypes of HPAI at the cost of \$0.8 million (CIDRAP, 2005). These vaccines were stored for five years as a preparedness measure against a potential outbreak.

Various reasons motivated such investments. First, the World Organisation for Animal Health (OIE, 2007) has recommended such stocks in preparedness for HPAI events. Second, a number of people in the community felt that adding emergency vaccination to the traditional stamping out strategy may help control the disease faster and reduce the risk of high economic damages. Third, in order to implement a vaccination strategy one needs to have an available stock, an ex-ante irreversible preparedness investment. Fourth, the incidence of HPAI was increasing in the world. Fifth, the United States has a large, valuable poultry industry which produced about 47.8 billion pounds of broilers, 7.1 billion pounds of turkeys, and 90.7 billion eggs in 2009.[2] An HPAI outbreak could be costly for both consumers and producers (Paarlberg et al., 2007).

Application of the balance model: a case study of HPAI in Texas

In this particular case study, the use of vaccination as a mitigation strategy requiring ex-ante investment in a stockpile is examined for the case of HPAI outbreak in the Texas poultry industry. The hypothetical outbreak in this analysis was assumed to start in a backyard flock of fewer than 400 birds. The investment cost for a vaccine stockpile was estimated to be approximately \$400,000 or 1.2 cents per dose (CIDRAP, 2005). The model explicitly incorporates an epidemiologic disease spread component where losses in combined consumers' and producers' welfare plus PP and RR costs are minimized with respect to HPAI mitigation strategies. Risk neutrality is assumed. Details about model specification and data can be found in Egbendewe-Mondzozo and others (2013).

The outbreak probability is uncertain. Consequently, the balance model is solved with a series of probabilities in an effort to determine how probable the outbreak needs to be in order to justify the investment in vaccine stock. We treat the probability of effective contacts between infectious flock of type (i) and a susceptible flock of type (j) as stochastic. The distributions used for inter-flock contact rates under various mitigation strategies were obtained using the results of surveys administered to poultry farm operators.

Monte Carlo simulation was used with 100 draws, and the results were tabulated into probabilistic distributions of the probability where the vaccination stock was justified. The obtained simulation results show that the threshold probability is sensitive to disease transmission parameters. The results show that 7 percent is the most likely trigger probability of occurrence that justifies the vaccine stocking investment. That is, over 100 simulations of the contact rates, the threshold probability of 7 percent occurs 82 times. Other higher threshold probabilities occur 18 times,[3] up to a maximum threshold probability of 31 percent that occurs with very low frequency. Some low probabilities of occurrence are also observed but with very low frequency.

Concluding comments

The decision of how much to invest ex-ante versus how much to rely on response and recovery ex-post is multifaceted and will vary widely based on disease, the location, and the susceptible population at that location. The balance model framework can be used conceptually and operationally to investigate the amount of ex-ante versus ex-post activity to undertake given estimates of disease probabilities and damage levels. The model can also be set up to find the levels of outbreak probabilities that justify alternative ex-ante investments. In turn, these breakeven probabilities can be presented to policy makers and/or experts who have better information pertaining to the likelihood of event occurrence.

The balance model can also be used more conceptually in thinking about responses to threats. For example, Elbakidze and McCarl (2006) indicate cases where ex-ante PP activity is more desirable: the event is more likely to occur, the PP costs are low, the response costs and potential damages are high especially if human life losses are incurred, the benefits of the PP investment are valuable in a number of settings beyond the target disease, the response cost and/or event damages loss reduction given the PP investment are high.

The balance model presented has several benefits and limitations. One benefit of the balance model is its flexibility in terms of required information about probability of occurrence or levels of damages. The model can be applied either in a way to estimate the probability of occurrence at which a particular level of investment is justified, looking over a range of investment possibilities, or the model can be solved over a range of probabilities and damage levels to estimate the level of investment that is justified. Elbakidze and others (2009) provide an illustration of this approach taking into account the implications of decision makers' risk preference structure in terms of designing optimal mitigation strategy.

The model is based on a point estimate of probability or damages and this is limiting. It can be extended with sensitivity analysis as above or be expanded to have a distribution of outbreaks of different severities. Naturally, the results are still only as good as the accuracy of the information going into the model. As presented, the array of PP and RR activities was narrow with a single investment portrayed. This approach can be relaxed with a continuum of investments and then have response activities conditional on them.

Based on the model and the case study presented here plus the studies we have done on this topic, it is difficult to make any broad conclusions about whether and when ex-ante PP action is preferred to ex-post response or vice versa. Actually, we think typically both are desirable but that the situation at hand can tip the balance between them. Nevertheless, we feel considering the problem in the framework of the balance model is useful in terms of generating qualitative and possibly quantitative insights into balancing PP and RR activities. This chapter has provided a framework for considering the balance between pre-event preparation and prevention activities and post-event response and recovery activities. Although simply presented, the framework can be expanded to capture the complexities associated with a spectrum of extreme disease events. As is the case with many frameworks and empirical models, the usefulness of the model for policy prescriptions depends on the extent to which available resources can be invested in tailoring and enriching the model with appropriate level of details for specific contexts.

Notes

1 For a full description see http://www.fao.org/emergencies/emergency-types/transboundary-animal-diseases/en/

2 Data retrieved from Quick Stats database: https://quickstats.nass.usda.gov/ (accessed July 8, 2017).

3 For examples of alternative presentation of results from models with stochastic disease spread parameters and implications for decision making in terms of optimal distribution of efforts across various mitigation options see Elbakidze and others (2009) and Egbendewe-Mondzozo and others (2013).

References

California Department of Food and Agriculture (CDFA). 2014. Exotic Newcastle disease—California historical reflection. Available at http://www.cdfa.ca.gov/ahfss/Animal_Health/newcastle_ disease_info.html (accessed April 2014).

Center for Disease Research and Policy (CIDRAP). 2005. USDA funds avian flu vaccine bank for poultry. Available at http://www.cidrap.umn.edu/cidrap/content/influenza/avianflu/news/nov0504vaccine.html (accessed June 26, 2017).

Centers for Disease Control (CDC). 2010. The 2009 H1N1 pandemic: Summary highlights, April 2009–April 2010. Available at http://www.cdc.gov/h1n1flu/cdcresponse.htm (accessed June 26, 2017).

Elbakidze, L. 2004. An economic exploration of prevention versus response in animal related bioterrorism decision making. PhD Dissertation, Texas A&M University.

Elbakidze, L. 2007. Economic benefits of animal tracing in the cattle production sector. *Journal of Agricultural and Resource Economics* 32(1):169–80.

Elbakidze, L., & McCarl, B.A. 2006. Animal disease pre-event preparedness versus post-event response: When is it economic to protect? *Journal of Agricultural and Applied Economics* 38(2):327–36.

Elbakidze, L., Highfield, L., Ward, M., McCarl, B., & Norby, B. 2009. Economic analysis of mitigation strategies for FMD introduction in highly concentrated animal feeding regions. *Review of Agricultural Economics* 31(4):931–50.

Egbendewe-Mondzozo, A., Elbakidze, L., McCarl, B., Ward, M.P., & Carey, J. 2013. Integrated economic-epidemic analysis of avian influenza mitigation options. *Agricultural Economics* 44(1):111–23.

Gramig, B.M., Barnett, B.J., Skees, J.R., & Black, J.R. 2006. Incentive compatibility in risk management of contagious livestock diseases. In *Economics of Livestock Disease Insurance: Concepts,* *Issues and International Case Studies*, edited by R. Koontz, D. L. Hoag, D. D. Thilmany, J. W. Green, and J. L. Grannis, 39–52. Cambridge, MA: CABI Publishing.

Gramig, B.M., Horan, R.D., & Wolf, C.A. 2009. Livestock disease indemnity design when moral hazard is followed by adverse selection. *American Journal of Agricultural Economics* 91(3):627–41.

Hagerman, A.D., McCarl, B.A., Carpenter, T.E., & O'Brien, J. 2012. Emergency vaccination and control of FMD: Consequences of its inclusion as a US policy option. *Journal of Applied Economic Perspectives and Policy* 34(1):119–46.

Hayes, D., Fabiosa, J., Elobeid, A., & Carriquiry, M. 2011. Economy wide impacts of a foreign animal disease in the United States. Working Paper 11-WP 525.

Jin, Y., McCarl, & B., Elbakidze, L. 2009. Risk assessment and management of animal disease-related biosecurity. *International Journal of Risk Assessment and Management* 12(2/3/4):186–203.

Kinsey, J., Seltzer, J., Ma, X., & Rush, J. 2011. Natural selection: 2006 E. coli recall of fresh spinach. *American Journal of Agricultural Economics* 93(2):629–35.

Lichtenberg, E. 2010. Economics of health risk assessment. *Annual Review of Resource Economics* 2:53–75.

Oladosu, G., Rose, A., & Lee, B. 2013. Economic impacts of potential foot and mouth disease agroterrorism: A general equilibrium analysis. *Bioterrorism and Biodefense* S12.

Paarlberg, P.L., Lee, J.G. 1998. Import restrictions in the presence of a health risk: An illustration using FMD. *American Journal of Agricultural Economics* 80(2):175–83.

Paarlberg, P.L., Seitzinger, A.H., & Lee, J.G. 2007. Economic impacts of regionalization of a highly pathogenic avian influenza outbreak in the U.S. *Journal of Agricultural and Applied Economics* 39:325–33.

Paarlberg, P.L., Seitzinger, A.H., Lee, J.G., & Mathews, K.H. 2008. *Economic impacts of foreign animal disease*. Economic Research Report Number 57. Washington, DC: U.S. Department of Agriculture Economic Research Service.

Pendell, D.L., Leatherman, J., Schroeder, T.C., & Alward, G.S. 2007. The economic impacts of a foot-and-mouth disease outbreak: A regional analysis. *Journal of Agricultural and Applied Economics* 39:19–33.

Rich, K.M., Winter-Nelson, A., & Miller, G.Y., 2005. Enhancing economic models for the analysis of animal disease. *Review of Science and Technology by the World Animal Health Organization* 24(3):847–56.

Shrestha, S.S., Swerdlow, D.L., Borse, R.H., Prabhu, V.S., Finelli, L., Atkins, C.Y., Owusu-Edusei, K., Bell, B., Mead, P.S., Biggerstaff, M., Brammer, L., Davidson, H., Jernigan, D., Jhung, M.A., Kamimoto, L.A., Merlin, T.L., Nowell, M., Redd, S.C., Reed, C., Schuchat, A., & Meltzer, M.I. 2011. Estimating the burden of the 2009 pandemic influenza A (H1N1) in the United States (April 2009–April 2010). *Clinical Infections Diseases* 52(Supp 1):S75–S82.

Torrey, E.F., & Yolken, R.H. 2005. *Beasts of the Earth: Animals, Humans and Disease*. New Brunswick, NJ: Rutgers University Press.

World Organisation for Animal Health (OIE). 2007. *Avian influenza vaccination: Information document—Verona Recommendations*. Available at http://www.oie.int/eng/info_ev/Other%20Files/A_Guidelines%20on%20AI%20vaccination.pdf (accessed June 26, 2017).

6 The feasibility and value of adaptive strategies for extreme risks

Robert Goble

Introduction

The ability to observe and adapt to changing circumstances is a desirable attribute for organizations that must respond to environmental and other sorts of threats. When there is substantial uncertainty about how events might develop over time and about how well particular response measures will work, flexible responses matched to changes as they are occurring are more likely to be effective. There is thus a strong case for attempting to put adaptive approaches into practice. "Adaptive management," a term drawn from the field of ecosystem management (Holling 1978; Walker et al. 2004), is by now used widely, and rather loosely, to describe a practical preparation for and reliance on this type of flexibility. There are two key aspects of this wider sense of adaptive management: (1) learning from experience, especially learning as situations develop over time, and (2) monitoring conditions and adjusting responses accordingly (Cox 2015; Smithson and Ben-Haim 2015). In planning to sail a boat from one island to another, you might want to learn in advance about the local shoreline and rocks, tides and weather patterns, and you certainly would not set a course without making provision for adjusting to the wind and currents that you will actually encounter.

Everyone can agree that learning and adaptability are goals worth striving for and, of course, since people are naturally resourceful, some learning and adjustment of response will occur as people attempt to cope with any hazardous event extreme or not (Furedi 2007). A recent news story observes that the successful Mexican response to the very intense hurricane Patricia was informed by previous experience, though it was also fortunate that the storm was localized and the storm surge was moderate (Neuman and Malkin 2015). The response to Sandy corrected mistakes made during Katrina (NOAA 2013) and there was a similar improvement in the response to Typhoon Hagupit a year after the disaster caused by Typhoon Haiyan (UNDP 2014). However, this learning, within organizations and in the broader society has been largely ad hoc. The concern in this paper is with adaptive management as a chosen strategy, a strategy that is planned and prepared for. The nuclear industry, for instance, has taken an adaptive approach to its routine, safety related operations (see Goble and Bier 2013 and references therein). But there has been little systematic planning for adaptation and learning to cope with extreme possibilities. While Guarnieri and Travadel (2014) have recently called for adaptive approaches to emergency nuclear plant operations (see also Golding et al. 1992, Goble and Hohenemser 1995, and Goble and Thompson 1995), they acknowledge that preparations are currently lacking and that there are important organizational obstacles. Good planning and preparation could have significant effects: various kinds of monitoring capabilities could and should be in place; better coordination among responders could facilitate adjustments to changing circumstances; and learning could be promoted by provisions for actively collecting and interpreting relevant information. It is worth remembering that while in general learning should be helpful, it is also possible to learn "wrong" lessons that might impede future coping. Explicit planning and preparation that facilitates learning could serve as a corrective to such tendencies.

The question we pose here is to what extent are planned and prepared adaptive strategies feasible and worthwhile for extreme risks, extreme in the sense that the threat is to overwhelm response preparations together with very large potential consequences? Several observations are pertinent.

1 Such planning and preparation are not prominent in the response plans for many environmental threats, both extreme and non-extreme. This suggests that there are broad-based institutional obstacles to the planned use of adaptation as a formal management strategy; these obstacles deserve attention.
2 Extreme risks appear to offer strong incentives for following adaptive strategies: there is high uncertainty about the course of potential events, so there will inevitably be considerable opportunity for flexible choices; and potential consequences can be large, so reductions in consequences would provide substantial benefits.
3 The nature of extreme risks, however, poses additional obstacles: extreme events are rare and idiosyncratic, so learning is a challenge (Lampel et al. 2009; van der Voort and de Bruijn 2009), particularly in a context where there will be much argument over blame and over broader policy questions (Dekker 2014); also, extreme events overwhelm response systems, and overwhelmed systems may have limited capabilities for acquiring information and for adapting based on that information.
4 When extreme events occur, it is often apparent that there has been insufficient effective planning and preparation of any sort: this could have been a matter of choice, if the event as it actually occurred was deemed too unlikely or too difficult to manage, or it could have resulted from institutional lapses; how much planning and preparation for extreme possibilities is worthwhile will always be an open question (Paté-Cornell 2004; Leveson 2011; Kenny 2015).
5 Most extreme events are extreme versions of not-so-extreme events for which some planning and preparation has been made; therefore, there will be value in planning to ensure that preparations for lesser events are useful rather than counterproductive in extreme circumstances.
6 The landscape for planning for extreme risks changes continuously: new knowledge and new technologies can help in making better preparations; political change, shifting populations, economic and infrastructure development, and an altered climate can create new vulnerabilities or improve resilience; new or transformed threats may appear.

These observations are drawn from descriptions of a broad range of extreme events of various sorts: hurricanes with floods (such as Katrina and Sandy, and recent typhoons in the Philippines) (Bier 2006; Daniels et al. 2006; GAO 2006; Blake et al 2011; Marshall 2013; NOAA 2013; Linkin 2014; UNDP 2014; Ayeb-Karlsson 2015); outbreaks of infectious diseases (Teh and Rubin 2010), including Ebola (Piot 2012; Farrar and Piot 2014; Pandey and Atkins 2014; Sack et al. 2014; WHO 2014; WHO 2015), AIDS (aids2031 2011; Piot 2012; UNAIDS 2015), and various flus (WHO 2011; Peiris et al. 2012; Rappuoli and Dormitzer 2012); earthquakes and tsunamis (Wyss 2004; Bilham and Gaur 2013; Tinti et al. 2015; Wyss 2015); nuclear accidents (including the combined tsunami and radioactive releases from the Fukushima nuclear complex) (OECD/NEA 2013; Guarnieri and Travadel 2014; Kingston 2014; Hobson 2015; Mosneaga 2015; Golding et al. 1992; Golding et al 1995; Goble 1995; Goble and Hohenemser 1995; Goble and Thompson 1995); and terrorist attacks (Kunreuther et al 2014; von Winterfeldt 2010). While specific characteristics of a particular extreme threat will matter crucially in detailed planning and preparation for specific threats, the observations above suggest that the question posed in this paper can be addressed in a general manner. Therefore, general lessons may be worth exploring as guides to more detailed planning. Further reasons for general exploration come from possible interactions between hazards: combined threats (such as the tsunami and subsequent nuclear accident) pose special challenges; also, many preparations for mitigating one type of threat are potentially useful in addressing other threats.

This paper reviews the above six observations in greater depth while discussing some steps for coping with various obstacles to adaptive management and demonstrating the general feasibility and usability of adaptive planning and preparation for extreme events. Along the way we will identify lessons that can also help in threat-specific planning.

In the next section we consider some characteristics of extreme risks and the challenges these characteristics pose to planning and preparations. Next, we look in more detail at various dimensions that distinguish extreme from less extreme threats. We develop the concept of a "planning frontier" to describe the level of attention to preparations for threats depending on their position with respect to these dimensions. In the following section we elaborate on the idea of a planning frontier to discuss ways of ensuring that efforts in preparation for lesser threats will be helpful if a more extreme event (on the other side of the planning frontier) should occur.

We next observe cases in which some of the harm ensuing from recent extreme events can be linked to complacency and lack of imagination in planning. This experience implies that vigilance is a needed complement to adaptive capabilities; we discuss the types of vigilance required. That discussion leads naturally to a section that considers the nature of resilience in the context of extreme risks in a changing environment. In the next to last section, we examine challenges and dilemmas that will arise in attempts to institutionalize vigilant and adaptive approaches. Finally, we summarize our answers to the initial question about the feasibility and usefulness of adaptive planning for extreme events and highlight the most salient lessons for the pursuit of adaptive approaches.

The spectrum of extreme risks I: general considerations

For the purposes of this paper, we do not need a sharp definition of "extreme risks." The two key characteristics are that extreme risks are threats to overwhelm available protective responses and they are threats to cause substantial harm (Bier et al. 1999). Such threats appear in a wide variety of forms: natural hazards such as storms and flooding, earthquakes and tsunamis; industrial accidents such as major accidents in nuclear or chemical plants; infectious disease outbreaks such as pandemic flu and Ebola. Slow-moving threats such as HIV/AIDS, persistent drought or, more broadly, global climate change, may still out-run response capability and qualify for the label "extreme." It is the relationship between the threat and response capability that is the focus of this discussion.

When response systems are overwhelmed it can be because they were inadequate to the task or because the available resources were used ineffectually. The response to Hurricane Katrina is an example of both (Bier 2006). Preparation is key both to providing response capability and to using it effectively. But preparations may be insufficient for several reasons: (1) surprise: the threat may not have been previously identified within the general practice of our sciences (a black swan) (Aven 2015, Chapter 2 this volume; Paté-Cornell 2012; Taleb 2007) or the threat may have been largely ignored for various reasons (including institutional failures, see item 3 below); (2) choice: the threat (in its extreme form) may have been considered too difficult, too expensive, and/or too unlikely to prepare for (a choice to prepare for a "100 year flood," but not a "500 year one" would be an example); (3) institutional failure: the responsible agencies might not have done their job, or responsibility might not have been clearly allocated (Bier 2006; Rubin 2012; Sack et al. 2014). These three reasons can, of course, overlap (Paté-Cornell 2004).

Reason (3) has been implicated in all too many extreme events; it, along with reason (2), should dominate the discussion of how to plan for extreme risks. Justifiable surprise is rare even when we only count extreme events, and some so-called surprises are really revelations of earlier wishful thinking. Fukushima surely was not a black swan. But "black swan" is not an empty concept. An example of the "black swan" sort of surprise is the emergence of the AIDS epidemic. No one anticipated the particular combination of the specialized interaction of the virus with the immune system, the very long period of dormancy, and transmission through sex and drugs so that social stress and denial amplified the dread associated with an unfamiliar disease. Indeed, it took several years to identify the virus and to characterize disease mechanisms and modes of transmission. And even now, 34 years after the first published identification of the disease (Gottlieb et al. 1981), we are still in the midst of the epidemic; inadequate infrastructure, the high cost of treatment, denial, stigma, and logistical obstacles combine so that more than twice as many people become newly infected for each new person receiving treatment for the disease (aids2031 2011; UNAIDS 2015).

In contrast, while the Ebola epidemic of 2014 also came as a surprise to most people, it was hardly a black swan. The old pattern of coping (through 22 previous outbreaks) already was risky and the risk was increasing. The disease and its virus were identified in 1976 (Piot 2012), and mainstream medical and public health science had preparations in place for such occurrences; the danger was obvious at the time, and there was a scramble by the international health community along with local authorities to control the outbreak. Partly because of measures learned and implemented on the fly while confronting the disease, and partly through luck (the disease failed to gain a footing in densely populated Kinshasa), the outbreak was contained and it then subsided (Piot 2012). Small outbreaks in isolated communities in several African countries occurred over the next 26 years and each of them was contained (WHO 2014). However, the medical systems in Guinea, Sierra Leone, and Liberia, were totally unprepared for the 2014 outbreak; facilities, supplies, even of the most basic sort, and trained health workers all were lacking (Pandey and Atkins 2014; Sack et al. 2014). And the international health community was not well prepared to intervene on the necessary scale. As of this writing, the epidemic appears finally under control, but the impacts around the globe still resonate. Part of the failure to prepare was complacency: previous outbreaks had all been small. However, the risk of a major epidemic, should the disease gain a foothold in a densely populated area, was obvious from the 1976 experience. And there was a further failure to consider and adapt to the changing conditions in a much more globalized world. In the decades after 1976, communities within a country have become much less isolated with much more travel back and forth. Equally significant has been the rapid increase in international travel. Health care practice has changed substantially, and that has demanded adaptations to limit transmission through health systems (Pandey and Atkins 2014; WHO 2014, 2015).

In a similar contrast, the appearance of the ozone hole and its connection to CFCs (Molina and Rowland 1974; WMO 1985) can be viewed as a black swan; however, climate change resulting from human carbon dioxide emissions should not be so viewed (unless we return to science before 1900 (Arrhenius 1896)).

Even though black swan extreme events are rare, having a non-empty "black swan" concept should help clarify analysis. Too much discussion about surprise, both within the risk community and between professionals and various stakeholders, has been about after-the-fact blame rather than about learning about different reasons for surprises and how to prepare better for them (Paté-Cornell 2012; Dekker 2014). To be able to make concrete comparisons about what is known and not known, to have a conversation about societal and institutional choices, and to consider various institutional difficulties and lapses can provide a contextually rich ground for such discussions. It thus can be useful to observe that a particular threat was or was not a black swan so that issues pertaining to choices and to institutional capabilities and competency must be addressed. Furthermore, an actual "black swan" sort of surprise poses its own challenges and these are worth considering on their own: the issues include (i) maintaining vigilance, so, as new science appears, its implications are assessed; (ii) developing response and recovery capabilities that can be useful even in unforeseen situations; (iii) maintaining learning capabilities that will be active when a black swan occurs.

Changing circumstances are implicated in many instances of preparation failure. Choices made earlier may well become obsolete because of changes in the conditions leading to the threat, changes in knowledge about the nature of the threats, changes in the technology available for coping, and changes in people's expectations. For instance, many dams and levees were sized decades ago based on expected run-off; development and construction since then may have increased run-off, while the dams have not been resized (FEMA 2013). Institutional and legal specifications can contribute to such problems. A man-made lake in Wisconsin was lost: while the dam that created the lake was monitored, shoreline erosion was not monitored since it was not "man-made," and the erosion led to a washout (Wikipedia 2015).

In the absence of corrective efforts, the rarity of particular types of extreme events makes such obsolescence more likely. For instance nuclear power emergency planning is still largely based on assumptions that were made 40 years ago about communications technology, about public use of communication technologies, and about plume tracking and predictive capabilities (Golding et al. 1992; Witt 2003). As noted above, changing patterns of travel and changes in medical services and protocols are implicated in the lack of planning for the recent Ebola outbreak. Capabilities for updating and adapting in preparing for response measures are thus a crucial aspect of adaptive approaches to extreme risks.

The spectrum of extreme risks II: dimensions of risk

Response measures may turn out to be insufficient in several ways. A storm or earthquake may be too intense, or it may affect too wide an area; it may happen too quickly or it may persist for too long. A disease may be too easily transmitted, or it may have an unusually long latency period. Complex interactions between threats may unravel preparations.

Planning and making preparations for potentially severe threats necessarily require making choices. The choices will involve commitments of resources, and it is important to keep in mind that human and institutional resources matter as well as economic ones, and that many resources are not easily exchangeable (Hattis and Goble 1994). It is desirable both that such choices be made explicit and that the planning process look beyond the boundaries of each choice. While the kinds of choices to be made will depend on the particularities of the hazard type, some generalization is feasible, and the generalizations can be illuminating. Most types of potentially severe threats have aspects that we might call intensity and likelihood: for hurricanes they are some measure of wind-speed (usually expressed as a "category" which ranges between 1 and 5) and storm frequency; for infectious disease they are some measure of infectivity and likelihood of emergence (perhaps from an animal reservoir); for an industrial accident such as a chemical plant or nuclear power plant accident they are a measure of the amount of toxic (radioactive) release and of accident probability. These measures may be combined: we speak of 100 year floods, for instance. However, combined measures should not obscure the nature of the choices, which involve choosing to make (or not make) particular practical preparations for a particular level of threat. Are we not making certain preparations because we think it is too unlikely we will need them? Or because we think that it is too difficult to prepare for a threat of such severity? Or, if it is a combination of those considerations, the nature of the combination should be made clear.

Severity and likelihood are not the only aspects that require choices in preparations. The spatial area that is under threat and the number of people who might be affected will be key considerations for planning and preparations. Likewise, temporal concerns will matter: how rapidly can we expect the event to develop? How long might it persist?

Qualitative features of a hazard also give rise to choices. Potential interactions and linkages between one hazard and another mean that we must choose the threat domain in making preparations. Do we prepare for earthquakes (and tsunamis) as part of nuclear or chemical emergency planning? And, conversely, do we consider complications from potential industrial accidents in preparations for earthquakes? Do we choose to integrate planning and preparations for different but possibly related threats? These can be organizational choices—who has what responsibility—as well as choices about whether or not a type of threat is considered.

Another important qualitative feature is the potential for intentional, possibly malevolent acts; acts of crime, sabotage, or terrorism. Is a concern for such possibilities made part of the planning process? Such acts present the added possibility of deliberate attempts at concealment and other measures to thwart preparations and responses; these possibilities create additional planning challenges. Furthermore, the public response may be different when there might be perpetrators of the hazard event: Will there be different demands for information? Will there be more incentive for panic? Will the task of locating perpetrators create additional demands?

It is not only possibilities for terrorism that affect public response. Some types of threats seem familiar while others, like releases of radioactivity or strange diseases, are unfamiliar and evoke dread. Do planners make preparations for addressing public fears as part of the overall effort?

We have listed a number of generalized aspects of threats that will require choices, either quantitative or qualitative, in the planning process. Key observations concerning them are:

1 These are distinct choices, thus distinct aspects that we may call dimensions of a threat.

2 They are described here in general terms for illustration, but can and should be defined specifically for each hazard and each planning process in question.
3 The characterization of the threat in terms of dimensions together with the set of planning choices define a planning frontier.
4 For events that may occur within that frontier, we can hope that we have chosen to make effective preparations, including a suitable allocation of institutional responsibilities.

The illustrative list of dimensions presented here is: intensity, likelihood, area at risk, population at risk, rapidity of occurrence, persistence, complexity/linkages, intentionality, and familiarity. It is interesting, but not surprising, that this selection, chosen by considering the practicalities of planning for the possible occurrence of various hazard events, overlaps considerably with the hazard dimensions identified in early attempts to categorize technological hazards in terms both of their physical character and public perceptions (Slovic et al. 1979; Hohenemser et al 1983; Kates et al. 1985; also see Slovic 2016). The significant findings in that work were that multiple dimensions were needed to make reasonable characterizations of types of hazards and that those dimensions also captured observed patterns in perceived risks.

Crossing a planning frontier

In normal planning and making preparations for hazardous events, people and institutions try to reduce the likelihood of those events and to reduce the consequences should the event occur. It is customary to consider planning for three phases of action; before-an-event actions directed toward reducing the likelihood or magnitude of potential consequences, the response during the event itself, should it occur, and the recovery after the event (Leonard and Howitt 2010). Supportive activities may include efforts to reduce surprises, to improve coordination among responsible parties, to avoid counterproductive responses and ineffective use of resources, to identify helpful response measures, and to allocate resources that are needed to enable and enhance suitable response measures. Good analytic tools are needed (Bier et al. 1999); they will partly be hazard specific, partly generic. They will be used in a context of choices, some explicit and some only implicit, of what to plan and prepare for. These choices collectively constitute the planning frontier (a related concept in systems engineering is that of a "performance envelope" or "competence envelope" (Wood 2006)).

An important class of extreme risks is the threat of possible events that lie outside the planning frontier. These are events that people have chosen not to prepare for or have chosen to make only partial preparations for; they might have deemed full preparations not feasible or too costly, or thought the events are too unlikely to matter. But, as was noted, the plans and preparations made within planning frontiers can still influence for good or ill the likelihood or severity of consequences of events outside the frontiers. We thus need a further goal for planning and preparations: to assure that the influence of preparations remains positive for events that lie beyond the frontier. Consideration of such influences must be built into the planning process: it is likely to require additional analytic tools.

For example, warning systems, in the event of a larger than anticipated flood or more intense storm, will not help if they direct people to a location that is vulnerable to the more intense event. Localized coordination among agencies may inadvertently develop barriers to coordination over a broader region. A classification system for directing response may not address extreme possibilities. Too much efficiency in the allocation of resources toward response measures may limit capabilities for responding to extremes. Or, put more generally, while it might not be considered cost-effective to fully prepare for events outside the planning frontier, some preparations might be cost-effective even if they are expected to be only partially helpful in extreme situations.

The ability to adjust and improve response preparations over time is critical. What is feasible and/or cost-effective may well change with changes in technology, changes in knowledge about threats, and changes in infrastructure and populations at risk. We already noted the importance of accounting for changing flood plain occupancy; a related concern is climate changes affecting storm frequency and intensity. Changes in medical practice and mobility were important in the Ebola epidemic; they can also affect many other health concerns. Changing communication technologies and changes in people's expectations about communication should influence the preparation and employment of warning systems; so should new capabilities for prediction and monitoring.

Effectively observing and adapting to such changes over time can take several forms: plans and preparations within a planning frontier will require updating; the question of whether plans and preparations will continue to be helpful for more extreme events must be revisited; and there must also be a capability for rethinking and, when appropriate, altering the planning frontier.

Vigilance is needed as a complement to adaptive approaches

Good planning is not enough: this is true even for threats that can be anticipated and are situated within a planning horizon, and holds even more strongly for extreme threats. Implementation matters. It is important to persist in paying close attention to possible warnings and in maintaining critical response capabilities. Ongoing vigilance is needed but there is a natural tendency of vigilance to "atrophy" as Freudenburg describes it (Freudenburg 1992). A "checklist mentality" can contribute to atrophy; this can be seen in the history of nuclear emergency planning (Golding et al. 1992, 1995; Lindbom et al. 2015)

There is a further and different need for ongoing attention. Conditions keep changing. Too much is unknown. Too much depends on unpredictable actions and interactions of individuals and institutions. There will be unanticipated difficulties with preparations and there may be unanticipated opportunities to be grasped. Complacency, a contributor to atrophy, will be a persistent danger and so too will lack of imagination. A focus on short-term threats and needs may neglect important long-term structural contributors to a hazard, or, conversely, too rigid a commitment to long-term goals may distract from immediate opportunities for improving conditions. Vigilance is also essential to address the surprises and the institutional failures that are not accessible to orderly adaptive planning.

Achieving the necessary vigilance in a changing and uncertain context for adaptation is difficult. The problem goes beyond the tendency to atrophy. Two types of vigilance are needed (Goble et al. 2016):

- Type-one vigilance: This is vigilance when you know what you are looking for: watching for warning signals; seeking to fill potentially troublesome gaps in knowledge; keeping up with predictable changes in the hazard context.

- Type-two vigilance: This is vigilance when you don't know what you should look for: preparing for and seeking surprises; discovering mistakes, misconceptions, undreamt of possibilities, and failures in institutional structures.

The dilemma is that the two types of vigilance are not compatible. The knowledge framework and focus that are essential to high competence at type-one efforts will interfere with type-two efforts and are likely to foment overconfidence. As an (extreme) illustration, the observer in a forest watchtower is looking for puffs of smoke; he or she is not likely to notice a gathering of arsonists or an invasion of moths.

Dilemmas such as this cannot be resolved. They set limits on what hazard management can achieve. However, even unresolved, they can be addressed as creative tensions. Further dilemmas arise in the process of institutionalizing adaptive approaches and incorporating vigilance. We come back to these dilemmas in discussing the institutionalization of adaptive approaches.

Aspects of resilience

Like many important descriptors of social systems and infrastructure, "resilience" is not easily defined; indeed it may be best to leave it undefined. The basic idea is to describe an ability to cope effectively with various stresses. There is thus considerable overlap between notions of resilience and of capabilities for hazard management. The distinction is primarily in emphasis. In discussing hazard management, the focus usually is on how to avoid or respond to a threat; while in discussing resilience, the focus is on the capabilities and effectiveness of the response system. Both kinds of discussion are needed, but they are insufficient in isolation. Resilience engineering offers a focus on the overlap: it is an argument that a resilient system capacity should be part of system design and maintenance (Hollnagel et al. 2006).

The term "resilience" evokes analogies: in materials engineering it is used to characterize a material's ability to return undamaged to its natural state after (repeated) flexes; in ecology its original use was to characterize an ecosystem capability for returning to equilibrium after a disruption (Holling 1978; but see also Walker et al. 2004 for an expanded characterization). Neither analogy suits very well the situations of concern for extreme risks and adaptive approaches to coping with them. The inevitability of a changing context means there is no natural equilibrium to return to. Moreover, the circumstances of an overwhelmed response capability mean that after the disruption of an extreme event, it is unlikely that a return to the pre-event state would be possible. Even if such a return were possible, it would not be desirable in most circumstances: changing contexts usually reflect dissatisfaction with the status quo. It would also be desirable if post-event conditions were less vulnerable to possible future threats. A better recovery is one that improves some conditions along with repairing damage, and one that is attentive to societal needs along with infrastructure (National Research Council 2011).

Even if the old analogies don't match, a notion of "resilience" to characterize desirable coping capabilities is useful in the context of extreme threats. Important capabilities include (Hollnagel et al. 2006; Kasperson 2008)

- *adaptability*: the ability to keep up with and adapt to changing circumstances and threats, changing populations with their needs and wants, and changing knowledge;

- *breadth of coverage*: capabilities for coping with a wide variety of circumstances, including some capability for responding to threats beyond the planning horizon;
- *flexibility in planning and response*: willingness to maintain incomplete but potentially useful preparations; capabilities for keeping track of conditions as they unfold; the ability to find and use resources as they become available;
- *attentiveness to recovery needs and opportunities*: planning and preparations for the recovery-needs created by extreme events; flexibility and resourcefulness to make societal improvements part of the recovery process.

In contrast to the homeostatic nature of the materials and ecosystem analogies, attention to these attributes considers resilience as a capability for embracing the dynamism and changing nature of human societies, for using the capacity to adapt to new and unfolding conditions as a resource, and for seeking positive outcomes even in adversity (Walker et al. 2004). Planning and preparations can make a difference.

Three dilemmas in institutionalizing adaptation and vigilance

The effectiveness of planning and preparations depends on the effectiveness of the implementing institutions. The institutions tasked with planning and preparing for a family of hazards must cope not only with the hazards, but also with a changing population and infrastructure, noisy debates, complex allocations of responsibility between institutions, and new knowledge and technologies. Adaptive approaches can be helpful, but they create particular challenges for these institutions, and part of the help is in identification of challenges. The challenges include

- to cover a broad range of possibilities across the spectrum of hazards within the family;
- to be capable of acquiring information and to use it flexibly in response;
- to have mechanisms for updating plans and resources to address changing circumstances;
- to assure that plans and resources will be helpful, not obstructive, for events that exceed the planning horizon;
- to maintain vigilance and reduce surprises.

Meeting these demands for flexibility, adaptability, and vigilance is made more difficult because of two dilemmas, one affecting internal arrangements within an institution, the other affecting the institution's external relationships (Goble et al. 2016).

The first, the internal dilemma, is how to maintain focus, coherence, and internal morale while persistently questioning assumptions and sharing responsibilities with other institutions. "We understand how to do this" and "We have always done things this way" are powerful bonds within an institution. So too is the need to defend the institution against external skepticism and challenges to its exercise of power. But such bonding discourages questioning and dissidence.

The second, the external dilemma, is how to maintain social trust and credibility in circumstances of high uncertainty, changing conditions, and divided responsibilities (Kasperson et al. 1992). Flexibility and waiting for more information can create an appearance of indecisiveness. Acknowledging uncertainty may win points for honesty, but it may raise questions about competence (Johnson and Slovic 1995; White and Johnson 2010).

These dilemmas and the third dilemma discussed earlier, the competing demands from two types of vigilance, cannot be resolved. At best, they will produce a creative tension within the institution, one that can only be maintained through a strong institutional culture. Effective and committed leadership is essential in instilling and reinforcing culture (Pandey and Moynihan 2006). Structural approaches can help as well. For instance, in many settings it may well be beneficial to have a separate department dedicated to the unforeseen prospects and questioning assumptions side of vigilance. Similarly, there will likely be benefits from formal institutional arrangements, a department or committees, dedicated to coordination with other involved institutions.

Extreme events in our usage are those that overwhelm immediate response capabilities. The direct threat may itself be overwhelming, but it is important to recognize that secondary or tertiary threats may also be serious; these pose additional institutional challenges since they may lie outside the expertise and authority of an agency focused on the direct threat. For instance, a response to pandemic flu might be complicated by issues pertaining to unemployment, housing evictions, utility shutoffs, childcare needs, needs for critical services, etc. (Bier et al. 2007). That means that in developing capabilities for response, realism about the possibility of disaster is called for, despite the institutional tensions that arise. The further implication is that planning, preparations, and resources will be needed for recovery efforts following the immediate emergency.

An institutional commitment to the maintenance and use of living risk assessments for the hazards in question (Goble and Bier 2013) could provide a formal mechanism for keeping track of needs and opportunities for updating planning. The living assessments would then provide a platform for incorporating new concerns, new knowledge, and for raising questions about assumptions. A complementary need is for serious efforts in planning and preparations for monitoring emergency conditions, keeping track of the response, and maintaining that monitoring through the recovery period. That work would be directed toward providing critical resources and intelligence in an actual emergency, but the process should also instill the idea of flexible responses and the importance of evaluation and learning.

Communication and information technologies are critical to emergency response efforts, but they are also key aspects of the context in which an emergency will play out. Changes in technology, in people's use of these technologies, and in people's expectations about the technologies have been profound over the past few decades, and rapid change continues. A responsible institution will dedicate resources and persistent attention both to evolving opportunities and challenges in its own communication, information gathering and assessment efforts, and also to assuring that those efforts are well-matched to the public's use of sources of information and communication technologies.

As is manifest in the second, external dilemma, institutions with responsibilities for emergency preparations and response have a responsibility for active engagement with the public. Such engagement is demanded by our democratic notions of governance: it is the public's welfare at stake. But engagement can have many practical effects as well: it may contribute to a more informed public that will respond better and can improvise when appropriate; it may bring out overlooked knowledge; public compliance or lack of compliance with directives may have significant consequences, so social trust and institutional credibility will matter (National Research Council 2011). These considerations will be especially important in the case of extreme threats: for these, there will be important and difficult choices about planning and the level of preparation; if an extreme event occurs and response capabilities are limited, credibility and social trust may be particularly valuable in reducing the amount of harm. In the context of public engagement, living risk assessments can have added utility as a forum for discussion with affected communities and interested parties.

Conclusions

The short answer to the question posed in the title is yes. In responding to a broad range of extreme threats, adaptive approaches are appropriate and feasible. They have the potential to reduce consequences substantially. However, recent experience with a variety of extreme events has shown that very often this potential is not fully realized in practice. Retrospective analyses have shown that active planning and preparation offer good prospects for improving responses and reducing consequences. The analyses show many missed opportunities, but there have also been examples of successes in learning and adaptation. The contrasting experience with Sandy after Katrina (NOAA 2013), and the sequence of Philippine typhoons, Hakupit following Haiyan (UNDP 2014), are two such examples. A more complicated, but illuminating example, is the ongoing evolution of responses to the AIDS epidemic (aids2031 2011; UNAIDS 2015).

The initial question was quite general and, in answering it, I considered extreme threats of many different sorts. Plans and preparations, if they are to actually function, must be sufficiently detailed to address threat-specific characteristics. However, the pathways toward improving plans can be described mostly in generic terms and so can many of the traps and detours that lead toward failure. Thus, a generic assessment of adaptive approaches is appropriate. A key conclusion from this assessment is that general efforts to encourage greater attention to planning and to more adaptive capability in emergency response would be valuable, and that the utility of good planning and preparation would extend into recovery.

Perhaps the most important current deficiency in adaptive capability is failure to learn from previous experience. In retrospective after retrospective following extreme events, one reads laments about how failures and errors are repeated. Misallocation of resources in Nepal follows a pattern similar to that in Haiti, for instance (Troutman 2015), and the failures of communication after Fukushima were remarkably similar to those after Three Mile Island and Chernobyl. Other reasons for pursuing adaptive strategies are also important. One, related to learning, is that the pursuit will encourage good monitoring and real-time assessment capabilities. Another reason is that attention to adaptation provides the best opportunity for taking account of changing circumstances and identifying and using new opportunities. Yet another reason is that adaptive approaches that look beyond the planning frontier will offer guidance to assure that measures designed for lesser events offer help rather than harm in the case of an extreme event.

Establishing and maintaining adaptive approaches is difficult, however. Adaptive approaches make special demands for learning, for gathering new information, and for changing responses over time. Vigilance, including persistent questioning of assumptions and expectations, is an essential component. If an institution is to effectively adopt an adaptive approach and to maintain a questioning vigilance, it will have to confront dilemmas that may well challenge the institution's identity.

We discussed three dilemmas that will constrain institutions in developing their adaptive capabilities: (1) an internal dilemma: how to maintain focus, coherence, and internal morale while persistently questioning assumptions and sharing responsibilities with other institutions; (2) an external dilemma: how to maintain social trust and credibility in circumstances of high uncertainty, changing conditions, and divided responsibilities; (3) a vigilance dilemma: how to combine vigilance when you know what you are looking for, with vigilance for what you don't expect.

Such dilemmas are inevitable in human institutions and cannot be fully resolved. However, they can be addressed constructively. Planning and preparation for more effective and more adaptive response to extreme threats will be fostered by

- broad institutional commitments to upgrading emergency response: keeping up with current circumstances and addressing governance issues (local, national, and international) merit particular attention;
- establishing dedicated offices within an institution devoted to vigilance and to coordination with other governmental and non-governmental institutions;
- explicitly characterizing the planning frontier in the planning process, and considering the implications for events that cannot be fully prepared for;
- attention to maintaining the availability of resources for recovery as well as for the immediate emergency;
- a substantially greater effort at providing resources and follow-up for monitoring and evaluation efforts;
- dedicated efforts to keep up with and adapt to changes in communication and information technology;
- enhanced efforts in public engagement.

In the aftermath of many recent extreme events we see some signs of learning and development of adaptive capabilities. Although the record is mixed, it provides further evidence for the feasibility of adaptive approaches, and it provides pointers for future improvements. It is worth emphasizing that we should expect that a successful development of adaptive strategies will itself follow an adaptive path, learning from past disasters, putting in place planning based on those lessons, and changing the planning as new evidence and knowledge is acquired.

References

aids2031 Consortium (2011) *AIDS: Taking a Long-Term View*. Upper Saddle River, NJ: FT Press.

Arrhenius, S. (1896) "The influence of carbonic acid in the air on the temperature on the ground," *The Philosophical Magazine and Journal of Science*, Series 5, 41(251), 237–276.

Aven, T. (2015) "Implications of black swans to the foundations and practice of risk assessment and management," *Reliability Engineering and System Safety* 134, 83–91.

Ayeb-Karlsson, S. (2015) "No need for luck to survive: why we should care about disaster risk reduction." United Nations University http://unu.edu/publications/articles/no- need-for-luck-to-survive-why-we-should-care-about-disaster-risk- reduction.html

Bier, V. (2006) "Katrina as a Bureaucratic Nightmare." In *On Risk and Disaster: Lessons from Hurricane Katrina*, edited by R.J. Daniels, D.F. Kettl, and H. Kunreuther, 243–252. Philadelphia, PA: University of Pennsylvania Press.

Bier, V., Haimes, Y., Lambert, J., Matalas, N., and Zimmerman, R. (1999) "A survey of approaches for assessing and managing the risks of extremes," *Risk Analysis* 19(1), 83–94.

Bier, V.M., Zach, L., King, S.B.J.D., O'Sullivan, T., and Burgos, I. (2007) "Decision Support for Pandemic Planning for the State of Wisconsin." Create Research Archive Non-published Research Reports. Paper 161.

Bilham, R. and Gaur, V. (2013) "Buildings as weapons of mass destruction," *Science* 341, 618–619.

Blake, E., Landsea, C., and Gibney, E., (2011) "The deadliest, costliest, and most intense United States tropical cyclones from 1851 to 2010 (and other frequently requested hurricane facts)," NOAA Technical Memorandum NWS NHC-6.

Cox, L. (2015) "Overcoming learning aversion in evaluating and managing uncertain risks," *Risk Analysis* 35(10), 1892–1910.

Daniels, R.J., Kettl, D.F., and Kunreuther, H. eds. (2006) *On Risk and Disaster: Lessons from Hurricane Katrina*, Philadelphia, PA: University of Pennsylvania Press.

Dekker, S.W.A. (2014) "The psychology of accident investigation: epistemological, preventive, moral and existential meaning-making," *Theoretical Issues in Ergonomic Science* http://dx.doi.org/10.1080/1463922X.2014.955554

Farrar, J.J., Piot, P., (2014) "The Ebola emergency—immediate action, ongoing strategy," *The New England Journal of Medicine* 371(16), 1545–1546.

Federal Emergency Management Agency (FEMA) (2013) "Flood Risks Nationwide." U.S. Department of Homeland Security. https://www.fema.gov/media-library-data/1381427654930-ab448ad691a3a8bc93f59c8f3b63fc49/FS_FloodRisksNationwide_092013.pdf

Freudenburg, W.R. (1992) "Nothing recedes like success? Risk analysis and the organizational amplification of risks," *Risk: Issues in Health and Safety* 3(1), 1–35.

Furedi, F. (2007) "The changing meaning of disaster," *Area* 39(4), 482–489.

Government Accountability Office (GAO) (2006), *Coast Guard: Observations on the Preparations, Response, and Recovery Missions Related to Hurricane Katrina,* GAO-06-903. Washington, DC: U.S. Government Accountability Office.

Goble, R. (1995) "What nuclear emergency planning can and can not accomplish." In *Preparing for Nuclear Power Plant Accidents: Selected Papers*, edited by D. Golding, J.X. Kasperson, and R.E. Kasperson (pp. 477–499). Boulder, CO: Westview Press.

Goble, R. and Bier, V. (2013) "Risk Assessment can be a game-changing information technology. But too often it isn't," *Risk Analysis* 33(11), 1942–1951.

Goble, R., Bier, V., and Renn, O. (2016) "Two types of vigilance are essential to effective hazard management: maintaining both together is difficult," Clark University (to be submitted for publication).

Goble, R. and Hohenemser, C. (1995) "Implications of the accident at Chernobyl for emergency planning." In *Preparing for Nuclear Power Plant Accidents: Selected Papers*, edited by D. Golding, J.X. Kasperson, and R.E. Kasperson (pp. 501–521). Boulder, CO: Westview Press.

Goble, R. and Thompson G. (1995) "The Use of Probabilistic Risk Assessment in Nuclear Emergency Planning." In *Preparing for Nuclear Power Plant Accidents: Selected Papers*, edited by D. Golding, J.X. Kasperson, and R.E. Kasperson (pp. 165–179). Boulder, CO: Westview Press.

Golding, D., Kasperson, J.X., Kasperson, R.E., Goble, R., Seley, J., Thompson, G. and Wolf, C.P. (1992) *Managing Nuclear Accidents: A Model Emergency Response Plan for Power Plants and Communities*. Boulder, CO: Westview Press.

Golding, D., J.X. Kasperson, and R.E. Kasperson, eds. (1995) *Preparing for Nuclear Power Plant Accidents: Selected Papers*. Boulder, CO: Westview Press.

Gottlieb, M., Schanker, H., Fan, P., Saxon, A., Weisman, J., and Pozalski, I. (1981) "*Pneumocystis cariini* pneumonia—Los Angeles," *Morbidity and Mortality Weekly Report* 30, 250–252.

Guarnieri, F. and Travadel, S. (2014) "Engineering thinking in emergency situations: A new nuclear safety concept," *Bulletin of the Atomic Scientists* 70(6), 79–86.

Hattis, D. and Goble, R. (1994) "Current priority-setting methodology: Too little rationality or too much?" In *Worst Things First,* edited by A. Finkel and D. Golding, 107–133. Washington, DC: Resources for the Future.

Hobson, C. (2015) "Rebuilding Trust after Fukushima." United Nations University Working Paper, UNU-IAS No. 4. United Nations University Institute for the Advanced Study of Sustainability.

Hohenemser, C., Kates, R.W., and Slovic, P. (1983) "The nature of technological hazard," *Science* 220(4595), 378–384.

Holling, C.S. (1978) *Adaptive Environmental Assessment and Management*. Chichester UK: Wiley.

Hollnagel, E., Wood, D., and Leveson, N. eds. (2006) *Resilience Engineering: Concepts and Precepts*. Farnham, UK: Ashgate.

Johnson, B. and Slovic, P. (1995) "Presenting uncertainty in health risk assessment: initial studies of its effects on risk perception and trust," *Risk Analysis* 15(3), 485–494.

Kasperson, R. (2008) "Coping with deep uncertainty: challenges for environmental assessment and decision-making." In *Uncertainty and Risk*, edited by G. Bammer and M. Smithson, 337–348. London: Earthscan.

Kasperson, R.E., Golding, D., and Tuler, S. (1992). "Social Distrust as a Factor in Siting Hazardous Facilities and Communicating Risks," *Journal of Social Issues* 48(4), 161–178.

Kates, R.W., Hohenemser, C. and Kasperson, J.X. (1985) *Perilous progress: managing the hazards of technology*, *Westview special studies in science, technology, and public policy*, Boulder, CO: Westview Press.

Kenny, K.E. (2015) "Blaming deadmen: causes, culprits, and chaos in accounting for technological accidents," *Science, Technology & Human Values* 40(4), 539–563.

Kingston, J. (2014) "Mismanaging risk and the Fukushima nuclear crisis." In *Human Security and Japan's Triple Disaster: Responding to the 2011 Earthquake, Tsunami and Fukushima Nuclear Crisis*, edited by P. Bacon, and C. Hobson, 39–58. Abingdon, UK, Routledge.

Kunreuther, H., Michel-Kerjan, E., Lewis, C., Muir-Wood, R., and Woo, G. (2014) *TRIA after 2014: Examining Risk-Sharing under Current and Alternative Designs*. Philadelphia, PA: Wharton Risk Management Center.

Lampel, J., Shamsie, J., and Shapira, Z. (2009) "Experiencing the improbable: rare events and organizational learning," *Organization Science* 20(5), 835–845.

Leonard, H. and Howitt, A. (2010) "Acting in time against disasters: a comprehensive risk management framework." In *Learning from Catastrophes,* edited by H. Kunreuther, and M. Useem, 18–40. Upper Saddle River, NJ: Pearson Education.

Leveson, N.G. (2011) "Applying systems thinking to analyze and learn from events," *Safety Science*, 49(1), 55–64.

Lindbom, H., Tehler, H., Eriksson, K., and Aven, T. (2015) "The capability concept: on how to define and describe capability in relation to risk, vulnerability and resilience," *Reliability Engineering & System Safety* 135, 45–54.

Linkin, M. (2014) *The Big One: The East Coast's USD 100 Billion Hurricane Event*. New York: Swiss Re.

Marshall, A. (2013) "Monster typhoon exposes an ill-prepared Philippines" Reuters, November 13. http://www.trust.org/item/20131113124938-bsv46/?source=spot…

Molina, M. and Rowland, F. (1974) "Stratospheric sink for chloroflouromethanes: chlorine atom-catalyzed destruction of ozone," *Nature* 239, 810–812

Mosneaga, A. (2015) *Tackling prolonged displacement: lessons on durable solutions from Fukushima*. Policy Brief No. 1. N.p.:United Nations University Institute for the Advanced Study of Sustainability.

National Research Council (2011) *Disaster Resilience: A National Imperative*. Washington, DC: National Academy Press.

Neuman, W. and Malkin, E. (2015) "Lessons of past disasters helped Mexico sidestep the brunt of a hurricane." *The New York Times*, October 24. http://www.nytimes.com/2015/10/25/world/americas/planning-helps-mexico-avoid-major-problems-from-hurricane-patricia.html

National Oceanic and Atmospheric Administration (NOAA) (2013) *Hurricane/Post-Tropical Cyclone Sandy October 22–29, 2012*. Service Assessment. Silver Spring, MD: U.S. Department of Commerce

Organisation for Economic Co-Operation and Development/Nuclear Energy Agency (OECD/NEA) (2013) *The Fukushima Daiichi Nuclear Power Plant Accident: OECD/NEA Nuclear Safety Response and Lessons Learned* NEA 7161. N.p.: OECD

Pandey, A., and Atkins, K. (2014) "Strategies for Controlling Ebola in West Africa." *Science Express*, October 30.

Pandey, S.K. and Moynihan, D.P. (2006) "Bureaucratic red tape and organizational performance: testing the moderating role of culture and political support." In *Public Service Performance*, edited by George A. Boyne, Kenneth. J. Meier, Laurence. J. O'Toole Jr., and Richard. M. Walker (pp. 130–151). Cambridge, UK: Cambridge University Press.

Paté-Cornell, E. (2004) "On signals, response, and risk mitigation." In *Accident Precursor Analysis and Management: Reducing Technological Risk through Diligence,* edited by J.P. Phimister, V. Bier, and H. Kunreuther, 45–60. Washington, DC: National Academy Press.

Paté-Cornell, E. (2012) "On black swans and perfect storms: risk analysis and management when statistics are not enough," *Risk Analysis* 32(11), 1823–1833.

Peiris, J., Poon, L., and Guan, Y. (2012) "Surveillance of animal influenza for pandemic preparedness," *Science* 335, 1173–1174.

Piot, P. (2012) *No Time to Lose: A Life in Pursuit of Deadly Viruses*. New York: W.W. Norton.

Rappuoli, R. and Dormitzer, P. (2012) "Influenza: options to improve pandemic preparation," *Science* 336, 1531–1533.

Rubin, C. ed. (2012) *Emergency Management: The American Experience 1900–2010*. Boca Raton, FL: CRC Press.

Sack, K., Fink, S., Belluck,P., and Nossiter, A. (2014) "How Ebola roared back." *The New York Times*, December 29. http://www.nytimes.com/2014/12/30/health/how-ebola-roared-back.html

Slovic, P., Fischhoff, B., and Lichtenstein, S. (1979): "Rating the Risks," *Environment* 21(4) 14–20, 36–39.

Slovic, P. (2016) "Understanding perceived risk: 1978–2015," *Environment* 58(1), 25–29.

Smithson, M. and Ben-Haim, Y. (2015) "Reasoned decision-making without math? Adaptability and robustness in response to surprise," *Risk Analysis* 35(10), 1911–1918.

Taleb, N.N. (2007) *The Black Swan: The Impact of the Highly Improbable*. New York: Random House.

Teh, J. and Rubin, H. (2010) "Dealing with pandemics: global security, risk analysis and science policy." In *Learning from Catastrophes*, edited by H. Kunreuther, and M. Useem, 211–234. Upper Saddle River, NJ: Pearson Education.

Tinti, S., Armigliota, A., Pagnoni, G., and Zaniboni, F. (2015) "Geoethical and social aspects of warning for low-frequency and large-impact events like tsunamis." In *Geoethics: Ethical Challenges and Case Studies in Earth Sciences*, edited by M. Wyss, and S. Peppoloni. 175–192 Amsterdam, the Netherlands: Elsevier.

Troutman, E. (2015) "What happened to the aid? Nepal earthquake response echoes Haiti" *AID. Works* June 19. http://aid.works/2015/06/nepal-haiti/?mkt_tok=3RkMMJWWf

UNAIDS (2015) *How AIDS Changed Everything*. Geneva, Switzerland: UNAIDS.

UNDP (2014) "Typhoon Hagupit: Preparedness measures have paid off in Eastern Visayas region of Philippines" December 10. http://www.asia-pacific.undp.org/content/rbap/en/home/presscenter/articles/2014/12/10/typhoon-hagupit-preparedness-measures-have-paid-off-in-eastern-visayas-region-of-philippines.html

van der Voort, H. and de Bruijn, H., (2009) "Learning from disasters," *IEEE Technology and Society Magazine 3*(28), 28–36.

von Winterfeldt, D. (2010) "Lessons from risk analysis: terrorism, natural disasters, and technological accidents." In *Learning from Catastrophes*, edited by H. Kunreuther and M. Useem, 171–189. Upper Saddle River, NJ: Pearson Education.

Walker, B., Holling, C.S., Carpenter, S.R., and Kinzig, A. (2004) "Resilience, adaptability and transformability in social–ecological systems." *Ecology and Society* 9(2) http://www.ecologyandsociety.org/vol9/iss2/art5

White, M. and Johnson, B (2010) "The intuitive detection theorist (IDT) model of trust in hazard managers," *Risk Analysis* 30(8), 1196–1209.

World Health Organization (WHO) (2011) *Strengthening Response to Pandemics and Other PublicHealth Emergencies: Report of the Review Committee on the Functioning of the International Health Regulations (2005) and on Pandemic Influenza (H1N1) 2009*. Geneva, Switzerland: WHO.

World Health Organization (2014) *Ebola and Marburg Virus Disease Epidemics: Preparedness, Alert, Control, and Evaluation*. Geneva, Switzerland: WHO.

World Health Organization (2015) *2015 Strategic Response Plan: West Africa Ebola Outbreak*. Geneva, Switzerland: WHO.

Wikipedia (2015) "Lake Delton." https://en.wikipedia.org/wiki/Lake_Delton#2008_washout Nov. 15, 2015

Witt, J. (2003) *Review of Emergency Preparedness of Areas Adjacent to Indian Point and Millstone*. Washington, DC: Witt Associates.

World Meteorological Organization (WMO) (1985) *Atmospheric Ozone 1985: Assessment of our Understanding of the Processes Controlling its Present Distribution and Change*. Geneva, Switzerland: WMO.

Wood, D. (2006) "Essential characteristics of resilience." In *Resilience Engineering: Concepts and Precepts*, edited by E. Hollnagel, D. Wood, and N. Leveson, 21–33. Farnham, UK: Ashgate.

Wyss, M. (2004) "Real-time prediction of casualties from earthquakes." In *Proceedings, Disasters and Society—From Hazard Assessment to Risk Reduction*, edited by D. Malzahn and T. Plapp, 165–173. International Conference, Universität Karlsruhe, Germany: Logos.

Wyss, M. (2015) "Do probabilistic seismic hazard maps address the need of the population?" In *Geoethics: Ethical Challenges and Case Studies in Earth Sciences*, edited by M. Wyss and S. Peppoloni, 239–249. Amsterdam, the Netherlands: Elsevier.

Part 3

Perceptions of extreme risks

7 It won't happen to me

The behavioral impact of extreme risks[1]

Eyal Ert and Ido Erev

Introduction

Many life activities imply a participation in a gamble that involves rare events. Flying to another country might (rarely) result in serious damage in the case that the plane crashes. Crossing a red light might also result in serious damage in the case of an accident. So might speeding, not using safety devices, or living in areas that are susceptible to floods, earthquakes, or volcano eruptions. Other examples include, but are not limited to, sending a paper for publication in a high-profile journal, investing money in new projects or in the stock market, buying products or services on the internet, engaging in unsafe sex, trying any medical drug that has some side effects, and the list goes on.

While the aforementioned activities are very different in many respects, they all share a common feature in that all of them involve rare events: outcomes (good or bad) that occur with a small probability. Another common feature of these activities is that the decision maker does not know a priori the exact probabilities of these rare events. Instead, the decision maker has to learn the outcome distribution over time relying on his own and/or other people's experiences.

The classical study of human decision-making tried to address the natural decisions exemplified above by the distinction between judgment (estimating probability) and decision processes. Implicit in this two-process approach is the assumption that people first estimate the probabilities of the feasible outcomes, and then choose between the different options by weighting the outcomes by their estimated probabilities (see Fox & Tversky, 1998). Most previous studies focused on one of the two processes. Basic studies of human judgment processes have compared intuitive probability estimates to objective probabilities. An exemplar of these studies (Phillips & Edwards, 1966) is presented in the first row of Table 7.1. The typical results of these judgment studies suggest that people tend to overestimate small probabilities. Experimental studies of choice behavior have examined one-shot choices in "decisions from description." The participants were asked to choose (once) between fully described payoff distributions. The typical results suggest that people tend to overweight the rare events. The second row in Table 7.1 summarizes one example of this observation.

Table 7.1 Summary of experimental studies that demonstrate overestimation of small probabilities and overweighting of rare events

Typical experimental task	*Typical results and interpretation*
Judgment (Phillips & Edwards, 1966) Urn A includes 30 Red Balls and 70 White balls. Urn B includes 70 Red Balls and 30 White balls. One of the two urns was randomly selected (the prior probability that A would be selected is 0.5). The experimenter sampled with replacement 4 balls from that urn. All 4 balls are Red. What is the probability that the selected urn is A?	**Overestimation** Mean estimate: 0.23 Bayes' posterior probability: 0.01 The mean estimate of 0.23 reflects overestimation of the objective small probability (0.01).
Choice (Kahneman & Tversky, 1979) Choose between the following two options: Option S: -5 with certainty Option R: -5000 with probability of 0.001; 0 otherwise	**Overweighting** Choice rate of option R: 20%. This choice rate suggests that most subjects behave as if the probability of the rare event (−5000) is over-weighted.

Source: Based on Marchiori, Di Guida, & Erev, 2015.

The coexistence of overestimation of small probabilities and overweighting of rare events appears to suggest that people are likely to exhibit extreme oversensitivity to rare events in decisions under uncertainty. Surprisingly, however, recent research shows that the exact opposite is often correct. Studies of decisions in situations in which people can use past experience to estimate the relevant probabilities reveal a bias toward underweighting of rare events (Barron & Erev, 2003; Hertwig et al., 2004). The typical decision maker behaves as if experience leads him or her to believe that "it won't happen to me." The findings from this recent research suggest a "description–experience gap" (Hertwig & Erev, 2009): People tend to overestimate the probability of rare events when they are asked to estimate them, and overweight rare events when they respond to a description of the potential risks. However, people tend to underweight rare events when they do not respond to descriptions of probabilities but rely instead on their own experiences.

The current chapter reviews, and tries to clarify, the study of experience on choice behavior. The first three sections summarize experimental studies that examine the robustness of the "it won't happen to me" effect. In the fourth section, we present new experiments that compare alternative explanations for the observed bias. The results suggest that experience reduces the weighting of rare events, but does not lead to complete neglect of these events. The last section is designed for readers interested mainly in the managerial implications of this phenomenon for risk management. This section presents empirical demonstrations of some implications of the "it won't happen to me" phenomenon to different domains of decisions under uncertainty that involve extreme risks.

Studies of repeated decisions from experience: the clicking paradigm

The clearest demonstrations of the "it won't happen to me" effect come from studies that used variants of the "clicking paradigm." The current section describes this paradigm and reviews the main experimental results.

The basic clicking paradigm

Barron and Erev (2003) used the basic "*clicking paradigm*" demonstrated in Table 7.2. The table presents its instructions and interface. The decision makers were asked to select between two unlabeled keys on the computer screen. Each click led to a random draw from the payoff distribution (realization of gamble) associated with the key that was clicked. The value drawn from the selected key determined the decision maker's payoff for that trial. The decision makers received no prior information concerning the relevant payoff distributions, and could only infer these distributions based on their obtained payoffs from previous choices.

Barron and Erev's studies show that immediate feedback tends to move behavior toward maximization, but there are robust exceptions. One exception was documented in the study summarized in Figure 7.1.

Table 7.2 The instructions and interface of a study that uses the basic clicking paradigm with partial feedback

Instructions	*Pre choice*	*Post choice*
The current experiment includes many trials. Your task, in each trial, is to click on one of the two keys presented on the screen. Each click will be followed by the presentation of your payoff from that click. Your final payoff will be determined by the payoff in one randomly selected trial.	Please select one of the two keys [ ] [ ]	Your payoff in this trial is 3 [3] [ ]

In the example the subject chose Left and won 3. The exact payoffs were determined by the problem's payoff rule. For example, in Problem 1, each Left choice leads to a payoff of 3, and each Right choice leads to a payoff of 32 in 10% of the trials, and 0 in the other 90% of the trials. The assignment of prospects to buttons and the order of the problems were randomly determined for each subject.

Source: Barron & Erev, 2003.

Problem	Option S	Option R
BE1	3	(32, 0.1; 0)
BE2	–3	(–32, 0.1; 0)

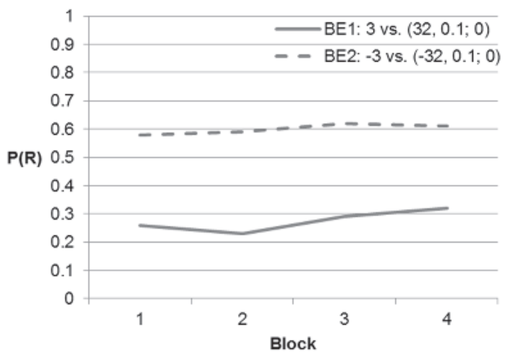


Figure 7.1 Problems BE1 and BE2 studied by Barron and Erev (2003). Problem BE1 involves a choice between 3 with certainty (Option S), and a gamble (32, 0.1; 0) that yields 32 with probability of 0.1 and 0 otherwise (i.e., with probability of 0.9; Option R). Problem BE2 involves choice between loss of −3 (Option S) and loss of −32 with probability 0.1. Subjects played each problem in the clicking paradigm for 400 trials. The figure shows the propotion of choices in the riskier Option R (P(R)) in 4 blocks of 100 trials each.

Both decision problems were played for 400 trials with the clicking paradigm, and the payoffs were in Israeli agorot (1 agora is 0.01 sheqel, and it was equal to about ¼ U.S. cent). In Problem BE1, one key (Option S) always led to a gain of 3, and the other key (Option R) led to a gain of 32 in 10% of the trials, and to a payoff of 0 in the other trials. The results revealed that the choice rate of the riskier option, which is associated with higher expected value (EV), (Option R, EV = 3.2), dropped with experience. The maximization rate over the 400 trials was only 28%. Thus, the subjects learned to prefer the safe option even though it is associated with lower expected value.

In Problem BE2, one key (Option S) always led to a loss of 3, and the other key (Option R) led to a loss of 32 in 10% of the trials, and to a payoff of 0 in the other trials. The results revealed that in this problem the subjects learned to prefer the riskier option. Since the riskier option (Option R) has lower expected value (−3.2), the results imply large deviation from maximization in this problem as well. Experience led the subjects to prefer the low expected value risky prospect.

The results observed in both decision problems are consistent with the "it won't happen to me" effect: the typical choice reflects low sensitivity to the rare but large outcomes (+32, and −32). Erev and Barron (2005) note that this low sensitivity can be a product of a tendency to rely on a small set of past experiences. For example, if the subject relies on a sample of four past experiences with each option from their memory in Problem BE2, she will select Option R 65.6% of the time (because 65.6% of the samples of size 4 do not include the realization of the rare event).

Clicking with full feedback: exposure to "forgone payoff"

Subjects in Barron and Erev's (2003) studies received "partial feedback": only their selected option was realized after each choice. Studies that relax this constraint and used the clicking paradigm with "full feedback" after each trial, which includes information on "forgone payoffs,"[2] show that the additional information increases the "it won't happen to me" effect (see Yechiam & Busemeyer, 2006). This enhancement is particularly clear when the rare outcome is unattractive (as in Problem BE2). It can be explained with the assertion that in these cases the tendency to underweight rare events is reduced by the hot stove effect, which refers to the asymmetrical influence of good and bad experiences on behavior[3] (Denrell & March, 2001; Fujikawa, 2009). The additional information given by the full feedback eliminates the hot stove effect.

Exposure to other people's decisions

A natural way of receiving information about "forgone payoffs" (i.e. payoffs from alternatives that were not chosen) in everyday life is by being exposed to other people's decisions. Yechiam et al. (Yechiam, Druyan, & Ert, 2008) compared decisions from experience with and without such social exposure. Specifically, they examined behavior in a typical clicking paradigm in one condition, and in another condition ("social exposure") subjects played the same task, but each subject could also see the actions and outcomes of another subject in real time. It was found that, similarly to the effect of forgone payoffs, being exposed to another person's decisions accelerates the "it won't happen to me" effect.

Clicking among multiple alternatives with full feedback

Although decision dilemmas often include more than two options to choose from, most studies have focused on binary choice for simplicity reasons. To examine the relation between the number of alternatives and risky choices, Ert and Erev (2007) extended the clicking paradigm with full feedback, by creating replicas of Options S and R. These replicas resulted in three different choice sets with 2, 6, and 50 alternatives to choose from. In one of the studies they examined a variant of Problem 1, and found that the proportions of choice in R were 23%, 39%, and 41% under the choice sets of 2, 6, and 50 alternatives respectively. This example shows that when the number of options in the choice set increases the "won't happen to me" effect might be somewhat reduced, though not eliminated.

The description-experience gap, and the joint effect of description and outcome feedback

The results reviewed above demonstrate the description–experience gap mentioned in the introduction: people exhibit oversensitivity to rare events when they decide based on the description of the incentive structure, and the opposite bias when they rely on past experience. In order to clarify the relative importance of the two biases, several studies used a variant of the clicking paradigm in which the subject sees the description of the payoff structure and in addition receives immediate feedback after each choice (Erev at al., 2017; Jessup, Bishara, & Busemeyer, 2008; Lejarraga & Gonzalez, 2011; Yechiam, Barron, & Erev, 2005). The results reveal an overweighting of the rare event in the very first trial, and a quick elimination or even reversal of this bias (i.e., underweighting of the rare event) with experience. Similarly, a description that is given after the subject has accumulated experience does not affect the "it won't happen to me" effect (Barron, Leider, & Stack, 2008). Together the results suggest that experience has a larger impact on behavior than description, and imply that warnings alone might not be sufficient to diminish the "it won't happen to me" regularity.

The coexistence of overestimation and underweighting of rare events

Barron and Yechiam (2009) studied a variant of problem BE2 (in which the loss from R was 20 with probability of 0.15), using the clicking paradigm with full feedback and one addition: starting at trial 201, the subjects were asked to estimate the probability of the rare outcome before they made their choice. The results revealed that the subjects overestimated the rare events, consistently with the examples presented in Table 7.1, but their choice reflected the opposite bias: they behaved consistently with the "it won't happen to me" effect.

Subjective evaluation of experience: the coexistence of the peak and freq. effects

The "it won't happen to me" effect appears to be inconsistent with studies that examine subjective evaluation of past experiences. This line of research reveals a "peak effect": it shows that, when evaluating (retrospectively and prospectively) unpleasant experiences, people tend to focus on relatively few extreme and rare key moments of the experience—the peaks (Fredrickson & Kahneman, 1993). In contrast, the "won't happen to me" pattern implies a "freq. effect": a tendency to pay more attention to the frequent experiences. Schurr, Rodensky, and Erev (2014) show that the two effects are not inconsistent; they can emerge in response to different questions. The requirement to evaluate past experience leads people to focus on the extreme and interesting cases, but the requirement to make repeated decisions leads them to exhibit the "it won't happen to me" effect.

The planning-ongoing gap

Another implication of the differences between subjective evaluation of probabilities and choice behavior relates to planning future behavior. This implication was illustrated by Yechiam, Barron, and Erev (2005). In one study they informed subjects that in each of the 100 trials of the experiment they will face the following choice problem:

Problem YBE (probability of Red is 1/200; payoffs in agorot, 100 agorot = \$0.25):
S: Loss of 8 if Red occurs, and loss of 2 otherwise
R: Loss of 100 if Red occurs, and loss of 1 otherwise

The subjects received a complete description of the payoff rule. Under condition "planning," the subjects were asked to plan their choices in advance. That is, they were asked to specify the number of trials in which they will play the risky option. Under condition "ongoing," the subjects had to specify their choice before each trial. The observed R-rate was 42% in condition "planning" (showing slight preference for the safer Option S), but 69% in condition "ongoing" (showing strong preference for the riskier Option R). That is, in line with one-shot decisions from description, planning decisions reflect higher sensitivity to rare events than ongoing decisions. Schurr et al. (2014) show that this effect of planning emerges even when planning is based on past experiences and the subjects cannot rely on a complete description of the incentive structure.

Studies of one-shot decisions from experience: the sampling paradigm

The cards sampling paradigm

Hertwig, Barron, Weber, and Erev (2004), and Weber, Shafir, and Blais (2004) suggested an alternative to the clicking paradigm to study decisions from experience: the "cards sampling" paradigm. In this paradigm the alternatives are presented as two decks of cards. The decision maker samples independent draws from each deck of cards as many times as she likes. Then, when she feels she had sampled enough from the decks, she can move on to a "decision stage" in which she has to choose once the deck from which she likes to draw a random card for "real money." Therefore, the most significant difference between the "clicking" and the "sampling" paradigms is that the latter separates between information acquisition (sampling stage) and choice (decision stage). This separation does not change the behavioral pattern: people still seem to behave as if "it won't happen to me." A recent study by Erev et al. (2010) compared the sampling and the clicking paradigms in 120 choice problems, and found that the level of risk-taking in the two paradigms is highly correlated ($r = 0.83$, $p < .0001$). However, this separation seems useful in studying the relation between information sampling processes and the decisions.

Representative samples

A typical observation in studies of the sampling paradigm is that people tend to be satisfied with small samples, typically 7–15 samples from each alternative (Hertwig et al., 2004; Hertwig & Erev, 2009). This tendency results with underrepresentation of rare events in the sampled payoff distribution and facilitates the "it won't happen to me" effect. Different approaches have been taken to explore the role of unrepresentative samples. Hau et al. (2008) forced subjects to draw 100 samples before they make their decisions to ensure they encounter the rare event. Ungemach et al. (2009; see also Camilleri & Newell, 2009; Hadar & Fox, 2009) fixed the sample so all payoffs are encountered during the sampling phase. The results show that while the effect of "it won't happen to me" is attenuated, it is still apparent even under representative samples.

Sampling multiple alternatives

Hills, Noguchi, and Gibbert (2013) examined the role of the choice set size in the sampling paradigm. They compared behavior in different conditions of the sampling paradigm that included 2, 4, 8, 16, and 32 alternatives. They found that when the number of alternatives increases, people sample more overall but sample each prospect less. Therefore, the increased number of alternatives facilitates the "it won't happen to me" effect.

Sampling ambiguous prospects and the promotion of new products

People tend to prefer "known risks" (whose probabilities are known) over equivalent "unknown risks" (whose probabilities are ambiguous), a phenomenon known as "ambiguity aversion" (Camerer & Weber, 1992; Ellsberg, 1961). Ert and Trautmann (2014) studied the effect of experience on ambiguity aversion by letting their subjects sample the "unknown risk." They found that this sampling experience eliminates the ambiguity aversion tendency, and leads people to exhibit the "it won't happen to me" effect. In a related study on product promotion, Ert, Raz, and Heiman (2016) studied consumer responses to products that are beneficial overall (i.e., whose expected values exceed their costs), but for which the distribution of values is highly skewed (e.g., longshot lotteries or safety-related products), so that they yield large benefits with low probability. In that context, they found that letting consumers experience these products before making a purchase decision was counterproductive, because consumers were too sensitive to the "typical" performance while experiencing the product (in which large benefits were most often not received). Therefore, such personal experience with a product actually lowered consumers' tendency to buy the sampled product despite its advantages.

"It won't happen to me" effect in other paradigms of decisions from experience

Most studies of decisions from experience have used variants of either the clicking paradigm, or the card sampling paradigm, reviewed above. However, some studies have used other paradigms and also found evidence for people exhibiting the "it won't happen to me" phenomenon. In this section we review some of these studies.

Probability matching

Most early studies of decisions from experience use the probability learning paradigm. In each trial of these studies (Estes & Suppes, 1959; Grant, Hake, & Hornseth, 1951; and see review by Vulkan, 2000) subjects have to guess which of two light bulbs will be turned on. Subjects do not receive prior information concerning the underlying probability, which is fixed throughout the experiment. To maximize performance the subjects should always select the bulb that was turned on more often in the past. The results reveal that the subjects deviate from this strategy in the direction of "probability matching." For example, Grant et al. (1951) studied a condition in which the optimal strategy led to accurate guess in 75% of the trials, and the results reveal that the choice rate of the optimal response was 75%.[4]

At first glance, the probability matching pattern appears to be inconsistent with the "it won't happen to me" effect; this deviation from maximization might indicate oversensitivity to the rare event. However, Erev and Barron (2005) show that the probability matching bias and the "it won't happen to me" effect can be products of the same cognitive tendency: the tendency to rely on small samples of past experiences.

Time saving decisions

Most studies of decisions from experience have focused on choice between monetary outcomes. Yet everyday choices are sometimes associated with nonmonetary outcomes, a typical example is the motivation of minimizing time delays (e.g., on daily commuting to work). Munichor, Erev, and Lotem (2006) replaced monetary outcomes with time delays (in seconds) in the clicking paradigm to find that people's timesaving decisions are similar to their money-related decisions from experience: a risky prospect is more attractive than a safer prospect with the same expected value only when it leads to a better outcome most of the time.

Signal detection tasks

Barkan, Zohar, and Erev (1998) examined choices in a signal detection task. In each trial subjects were presented with a square placed at a certain height on a screen, and had to classify that square (the signal) as "High" or "Low" on their screen. In one condition, the decision "Low" maintained the status quo, and the decision "High" could lead to a gain of 1 point if "High" was the right classification, or a loss of 10 points if it was the wrong one. The exact probability of a gain increased with the height of the square, and the optimal strategy implies a cutoff rule, i.e., decide "High" only when the square/signal exceeds a certain height on the screen. The results reveal that most subjects deviated from the optimal cutoff in the direction of the "it won't happen to me" effect. For example, they mostly selected "High" (i.e., their cutoff was lower than optimal), even when the probability of the large loss was 10%–20%. Notice that in these cases the signal's true classification is only rarely "Low," so selecting "High" leads to a better outcome most of the time. However, such decisions also imply a negative expected return since the expected loss is higher than 1, while the expected gain is less than 0.9 in these cases.

Other animals

Shafir et al. (2008) compared repeated choices of humans and honeybees (*Apis mellifera*) between a "safe" option that provided a gain with certainty, and a risky option that provided a higher gain but was associated with a rare relative loss. The results showed that when those options could be easily discriminated both humans and honeybees did not pay enough attention to the possibility of the rare relative loss. Both species behaved as if "it won't happen" to them.

Interpretation and boundary conditions

Kahneman and Tversky (1979) hypothesized that "because people are limited in their ability to comprehend and evaluate extreme probabilities, highly unlikely events are either neglected or over-weighted, and the difference between high probability and certainty is either neglected or exaggerated" (p. 283). Under one interpretation of the results summarized in the previous sections, they clarify the conditions under which rare events are neglected: Neglect is more likely in decisions from experience than in decisions from description. Under a weaker interpretation of the results, the "it won't happen to me" effect does not reflect a complete neglect of the rare events. Rather, it implies that experience reduces the subjective weighting of the rare events. The experiments described below compare the two interpretations.

Rare events: neglected or underweighted

The main difference between the "neglect" and the "underweighting" explanation of the "it won't happen" effect involves the assumed effect of the expected payoffs. If rare events are neglected, their effect on the expected payoff is irrelevant. In contrast, underweighting implies sensitivity to the effect of the rare event on the expected payoffs; extreme rare events can drive choice behavior even if they are underweighted. We investigate this by testing whether subjects are more attentive to rare events when their consequences are sufficiently severe (as in the "Distinct EV" choice problem presented in Figure 7.2) than when they are less severe (as in the "Similar EV" problem presented in Figure 7.2).

Both problems involve a choice between a certain payoff (2.52), and a risky prospect that usually pays a higher payoff (2.53), but pays a lower payoff with low probability. The problems differ with respect to the outcome of the low probability event. In Problem Similar EV the outcome of the low probability event is moderate (its distance from the common outcome is only 10 times the difference between the common outcome and that of choosing S), and in Problem Distinct EV the rare outcome is significantly lower, implying a noticeable difference between the EVs of the two options. The assertion that people neglect/ignore rare events implies a high preference for R in both problems, since Option R is better in 89% of the time, and EVs would be similar if the rare outcome is ignored. If, however, the outcome of the rare event is not neglected but underweighted, then, even though Option R looks more attractive most of the time, people may learn to notice the significant difference in EV that is implied by this option. Thus this hypothesis implies a preference for R in Problem Similar EV, but not in Problem Distinct EV where the rare event has higher impact on the EV of Option R.

Problem	Option S	Option R
Similar EV	2.52	(2.53, 0.89; 2.43) [EV = 2.519]
Distinct EV	2.52	(2.53, 0.89; 2.03) [EV = 2.475]

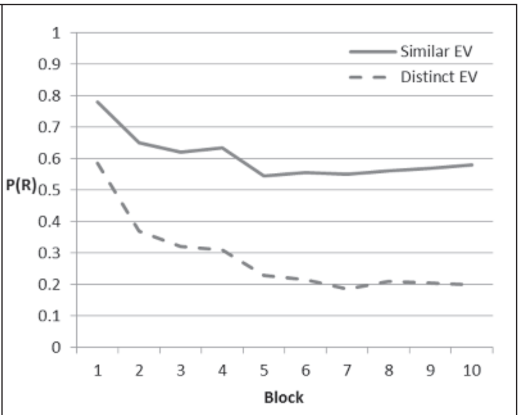


Figure 7.2 Proportion of choices in Option R (P(R)) in 10 blocks of 40 trials in Problems Similar EV (SEV) and Distinct EV (DEV). The notation (x, p; y) refers to a prospect that yields outcome x with probability p and outcome y otherwise (with probability 1−p).

In order to compare the two interpretations, we let 40 Technion students play the clicking paradigm with full feedback for 400 trials and for real money. Half of them played Problem Similar EV and the other half played Problem Distinct EV.

The results reveal that subjects learned to avoid the risky option when the rare event had high impact on their expected value (red line in Figure 7.2, but much less so when the rare event had low impact on their expected value (blue line in Figure 7.2). More specifically, the rate of risky choice (i.e., choice in Option R) across the 400 trials was 60% in Problem Similar EV, but only 28% in Problem Distinct EV. Figure 7.2 shows that the first block of trials reveals high R-rates, which did not significantly differ between problems: 80% in Problem Similar EV and 69% in Problem Distinct EV. With time the mean R-rates mildly reduced in Problem Similar EV, by 21% from the first to the last block, and significantly reduced by 48% in Problem Distinct EV.[5]

Together the results support the hypothesis that rare events are underweighted, not neglected: subjects' choices reflected insufficient attention to rare events, but when rare events were sufficiently important subjects learned to take them into account.

Is underweighting affected by absolute or relative EV differences?

The finding that rare events are underweighted, rather than neglected, raises a natural question: exactly what conditions would make the outcome of the rare event "sufficiently significant" to impact behavior? One account for the current findings is that people attend to the *absolute* EV difference between the proposed options. Another account suggests that people respond to the differences in expected value *relative* to the payoff variance (Weber et al., 2004).

We conducted an additional study to differentiate the "sensitivity to absolute EV" from the "sensitivity to relative EV" hypotheses. The study compares the problems presented in Figure 7.3. In all problems, Option R is associated with lower EV than Option S. Yet in Problem ADEV (Absolute Difference in EV) the difference between the expected values is 11 times higher than in Problem Base. Therefore, in accordance with the results

Prob.	Option S	Option R	EV diff.	Relative EV diff.[1]
Base	0	(1, 0.9; -10)	–0.10	–.030
ADEV	0	(1, 0.9; -20)	–1.10	–.175
RDEV	0	(0.1, 0.9; -2)	–0.11	–.175

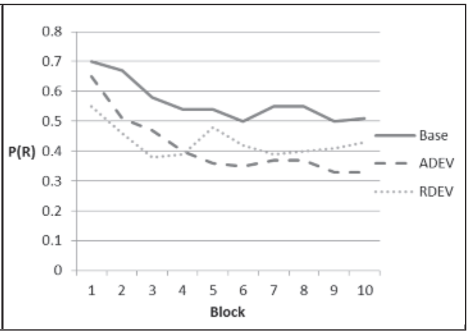


Figure 7.3 Proportion of choices in Option R in 10 blocks of 40 trials in Problems Base, ADEV (Absolute difference in EV), and RDEV (Relative difference in EV). Each problem involves a choice between Option S that yields 0 with certainty, and Option R that yields x with probability of 0.9, −y otherwise, when x equals 1,1, and 0.1 and y equals −0.10, −1.10, and −0.11 in Problems Base, ADEV, and RDEV respectively. Payoffs were in sheqels, 1 sheqel = $0.25. The notation (x, p; y) refers to a prospect that yields outcome x with probability p and outcome y otherwise (with probability 1−p).

reported in the previous section, we predict that Option R will be more attractive in Problem Base than in Problem ADEV. Problem RDEV (Relative Difference in Expected Value) is added to examine the role of the relative EV differences. The new Problem RDEV was designed so that it is similar to Problem Base with regard to the absolute EV difference between the safe and risky options, but similar to Problem ADEV with regard to the relative EV difference. Thus, if the absolute EV difference drives the effect of rare events, then the R-Rate in Problem ADEV is expected to be lower than the R-rate in the two other problems. Alternatively, if relative difference in EV matters then the R-rates in both Problems ADEV and RDEV are expected to be similar, but lower from the R-rate in Problem Base.

Forty-eight Technion students participated in this study. The procedure was identical to that of the study reported in the last section, with the exception that each subject played all problems, which were presented in a random order. Each problem was played for 100 trials.[6]

The results reveal that the R-rate (i.e., proportion of risky choice) in Problem ADEV (41%) is significantly lower than in Problem Base (56%). This result replicates the findings from the study in the previous section. The Overall R-rate in Problem RDEV (43%) is also significantly lower than in Problem Base.[7] However, no significant difference in R-rate was found between Problems ADEV and RDEV.[8] Thus, it seems that the difference in EV *relative* to the payoff variance (rather than the absolute EV difference) is the factor that accounts for the observed behavioral differences across problems.

Two variants of a basic sampler model

Previous research has shown that the basic properties of decisions from experience can be captured with simple models that assume noisy reliance on past experiences (Erev, Ert, & Yechiam, 2008). The basic model assumes random choice in the first trial, and that each of the following decisions is based on a small sample of k_i (property of agent *i*), randomly drawn past experiences. The agent selects the risk option if and only if:

(1) $SMR - SMS > 0$

where SMR, and SMS are the mean payoffs in the sample of K draws from R and S respectively. The sample size K_i is a free parameter that reflects a property of the different agents and is drawn from the set {1, 2,… K}. The model best captures the data of the two studies summarized in the two preceding sections, with an estimated parameter of K_i = 23, the Mean Square Deviation (MSD) is 0.003.

The "sensitive sampler" model is a variant of the basic model that assumes sensitivity to the grand mean. It is identical to the basic model with the exception that equation 1 is replaced with:

(2) $(GMR - GMS) + (SMR - SMS)$

where the added terms GMR and GMS are "grand means," i.e., the mean payoffs from R and S respectively in all previous trials. The sensitive sampler model best fits the data with an estimated parameter of K_i = 17, and has an MSD of 0.0009.

Comparison of the two models shows that the added assumption of the agent's sensitivity to the alternative's grand mean improves the model's descriptive value without any addition of parameters. The better fit results mostly from the model's higher sensitivity to the differences between the problems' EV, such as the difference between Problems Similar EV and Distinct EV. Notice that the basic sampler model already captures some of this sensitivity by assuming that the agent's sample size is relatively large. Accordingly, the sample size is estimated to be much lower in the sensitive sample model (K_i = 17) than in the basic sampler model (K_i = 23).

The impact of extreme risks and potential implications

The experimental studies summarized above focus on abstract situations, associated with relatively small objective risk. The choice task involved clicking, and the maximum loss from risky behavior was less than $3. The focus on abstract situations in the lab allows the study of a fully controlled environment that is crucial to differentiate hypotheses (as demonstrated in the section above). In the current section, we overview several field studies that evaluate the potential implications of the "it won't happen to me" phenomenon in the "real world."

Car radio with detachable panels

The difference between planning and ongoing decisions suggests that sometimes people buy safety devices, but "learn" with time to neglect them. One example of this buying–(not)using gap is a study, by Yechiam et al. (2006), that focuses on car radios with a detachable panel that were popular in Israel when this study was conducted. The detachable radio panel provides protection against theft but only when it is used (i.e., when the panel is detached).

The decision to buy a detachable panel relies on its value as a safety device. The decision not to detach the panel is made without an explicit presentation of a threat, and is mostly shaped by experience. Thus, the "it won't happen" effect implies a decrease in the tendency to use the panel with experience, as the small probability of theft is underweighted. Yechiam et al. surveyed Israelis on this issue, and found that the large majority (96%) of Israelis who bought car radios between 1995 and 2003 preferred the type with a detachable panel although it was more expensive. Most respondents detached the panel in the first two weeks, but were much less likely to detach it after a year. That is, responders seemed to underweight the probability of theft when they made their decisions from experience, despite their initial concern of theft that led them to buy such expensive detachable panels in the first place.

The effect of rare terrorist attacks

The impact of terrorism on tourism represents one of the main economic effects of terrorism (Weimann, 1993). The "it won't happen" effect may have several implications: first, it suggests that the frequency of attacks might be more effective than their magnitude. A study conducted on the impact of terrorism on tourism demand in Israel during 1991–2001 (Pizam & Fleischer, 2002) found just that: the frequency of acts of terrorism had caused a larger decline in international tourist arrivals than the severity of these acts. The second implication is that local residents would be less impacted than international tourists as the former have gathered safe experience, and underweight the rare event of a terrorist act (Yechiam et al., 2005).

The enforcement of safety rules

The research reviewed above has three implications for the design of safe working environments (see Erev & Rodensky, 2004; Schurr, Rodensky, & Erev, 2014; and related ideas in Zohar, 1980). First, the results suggest that rule enforcement is necessary even when safe behaviors, e.g., the use of safety equipment, are the optimal course of action. An explanation of the relevant risks might not be enough, as when workers make decisions from experience they are likely to underweight the low-probability-high-hazard event and behave as if they believe that it won't happen to them.

A second implication concerns the effectiveness of rule enforcement systems in which a small proportion of the violations are severely punished (Becker, 1968). Systems of this type are likely to be effective in the context of decisions from description, but less effective in contexts of decisions from experience since low-probability punishments are likely to be underweighted. The third implication suggests that the fact that workers take unnecessary risks and seem to ignore safety rules does not imply that they would object to attempts to enforce these rules. Indeed, the planning–ongoing gap implies that when workers are explicitly asked to consider the safety issue they would agree that they want to behave safely, and would be happy to see that the management designs a rule enforcement system to help them achieve this goal.

Finally, the arguments presented above suggest that behavior is much more sensitive to the probability than to the magnitude of the punishment. Thus, a gentle Continuous Punishment ("gentle COP") policy that implies low punishments with high probability can be very effective as long as the fine is larger than the benefit from violations of the rule and the risk of avoidance behavior is low.

Erev and Rodensky (2004, and see Schurr et al., 2014; Zohar & Erev, 2007) applied this "gentle COP" method in 12 Israeli factories. They designed a mechanism by which supervisors approached each worker who violated a safety rule and reminded him that this behavior might result in injury, and would be recorded if repeated. The official role of these "violations records" was to allow the management to positively reinforce workers who followed the safety rule by giving these workers a higher probability of winning a lottery. Baseline data were collected about two months prior to intervention, which included objective measures of the workers' safety behaviors. Figure 7.4 presents measures of adherence to safety rules before and during the intervention in one department of one of these factories. The data was collected by the research team, and was independent of the supervisors' comments and records.

As the example in Figure 7.4 shows, the intervention had a large and immediate positive effect. A similar pattern was observed in all 12 factories. The rate of safe behaviors increased to 90% immediately after the beginning of the intervention and did not diminish within the next two years of measurement. Given the success of the intervention, and its relatively low cost, the factories have decided to maintain the experimental policy.

The limitations of warnings

Warnings are considered to be an important information source that is designed to protect people by encouraging them to avoid risks. People's tendency to overweight the probability of rare disasters when they respond to described information suggests that warnings can be indeed very effective in facilitating safety. Nevertheless, the "it won't happen to me" effect suggests important limitations to the effectiveness of warnings. One limitation suggests that when people are warned against engaging in risky behavior only after having past experience, the warning is ineffective in eliminating the "it won't happen to me" effect (see Barron et al., 2008). Another limitation refers to cases where the accuracy of warning information is uncertain (e.g., weather warnings). In such cases

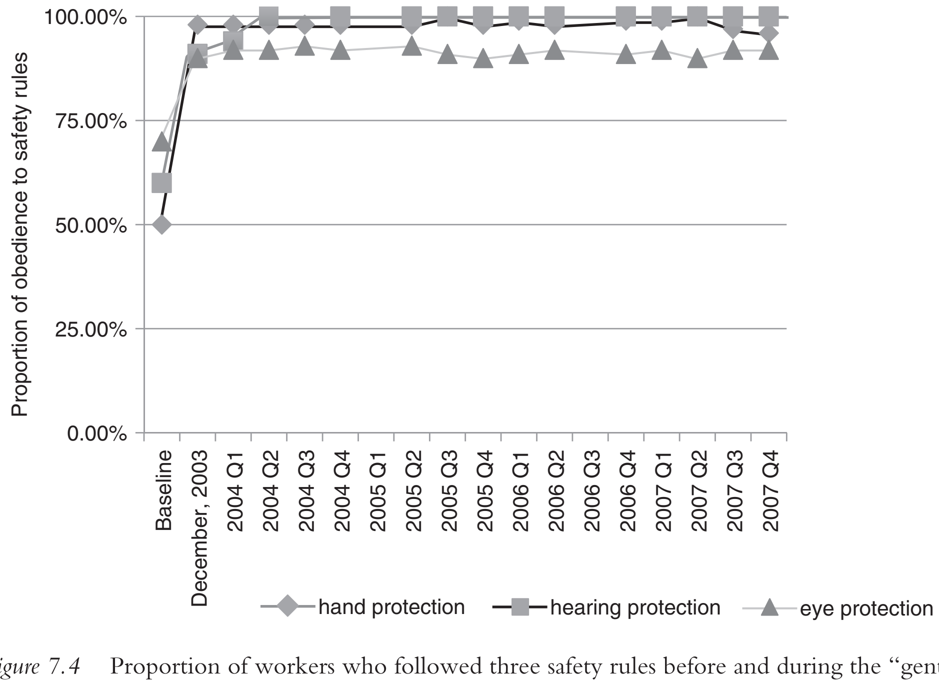


Figure 7.4 Proportion of workers who followed three safety rules before and during the "gentle COP" intervention (reproduced with permission from Schurr et al., 2014).

a tradeoff exists between the probability of detection and false alarms. Intuition suggests that the costs of false alarms of rare disasters (e.g., tornados) should be much lower than the costs of failing to take heed of them. Yet experiencing false alarms might lead people to learn erroneously that the warning is ineffective and discourage them from taking the necessary safety measures. In line with this concern, a recent examination of the relation between false alarm ratios and tornado casualties across the U.S. from 1984–2004 revealed that one standard deviation in the false alarm ratio increased expected fatalities by between 12% and 29% (Simmons & Sutter, 2009).

Moral hazards

Insurers and economists are often worried about the possibility that having insurance might change behavior of insured persons in ways that will increase insurance claims. Moral hazard is defined as the "intangible loss-producing propensities of the individual assured" (Pauly, 1968). Economists have differentiated between "ex-ante" moral hazard, which refers to taking more risks (e.g., reduction in preventive effort) once the individual is insured, and "ex-post" moral hazard, which refers to increasing demand for medical care. Zweifel and Manning (2000) review the evidence for both types of moral hazard and find that, at least in health-care, the evidence for ex-ante moral hazard is weak, while the evidence for ex-post moral hazard is much stronger. The "it won't happen to me" phenomenon can explain this interesting difference. It suggests that people take risks because they believe that "it won't happen to me" and underweight the probability of rare event, and they are much less affected by its magnitude. Thus, reducing the outcome magnitude of the rare event is not likely to change risky behavior.

Summary

The current analysis suggests that there are many situations in which the typical decision maker behaves "as if" he or she believes that "it won't happen to me." The tendency to underweight rare events is particularly robust when people make ongoing decisions based on past experience, and the difference between the expected values of the different options is relatively small. These results can be captured with the assumption that decisions from experience reflect a tendency to rely on a small sample of past experiences. In addition, the results show high sensitivity to rare events in one-shot decisions from experience, in planning decisions and in probability estimation. Review of field and intervention studies suggests that the understanding of the conditions that trigger under- and overweighting of rare events can be used to design policies that facilitate wise reaction to extreme risks.

Notes

1 This research was supported by grants from the Israel Science Foundation.
2 "Full feedback" includes the realization of the obtained payoff, but also the information about the "forgone" payoff from the key that was not selected in that trial.
3 Specifically, a bad experience with the risky alternative decreases the likelihood of exploring that alternative again, thus preventing future good experiences with that alternative that could correct the bad impression.

4 Follow-up research (see Bereby-Meyer & Erev, 1998; Edwards, 1961; Siegel, 1961) shows that the magnitude of the deviation from optimal choice is diminished with time in longer experiments. Yet, the bias in the direction of probability matching does not disappear.

5 The difference in R-rate between Problems Similar EV and Distinct EV is significant, $t(38) = 3.02$, $p = .004$; the reduction between the first and last blocks in Problem Distinct EV is significant, $t(19) = 5.5$, $p < .0001$.

6 Problem Base is reported by Nevo and Erev (2012), and the other two problems were added for this study. Subjects were also presented with a fourth "filler" problem (a choice between a sure 0 and a gamble that yields 1 with probability of 0.1 and −10 otherwise) that was run for the purposes of another study.

7 The difference between R-rates in Problems ADEV and Base is significant, $t(47) = -4.03$, $p < .001$; R-rate in problem RDEV also significantly differs from Problem Base, $t(47) = 2.76$, $p = .008$.

8 Figure 7.3 presents the R-rate in each of the problems over 10 blocks of 10 trials. The figure reveals that the R-rates were high in all conditions in the first block of trials (70%, 65%, and 55% in Problems Base, ADEV, and RDEV respectively). It also shows significant reductions in risk taking from the first to the last block in all conditions: 19%, $t(47) = 3.75$, $p = .0005$. in Problem Base; 32%, $t(47) = 6.01$, $p < .0001$ in Problem ADEV; and 11% $t(47) = 1.91$, $p = .063$ in Problem RDEV.

References

Barkan, R., Zohar, D., & Erev, I. (1998). Accidents and decision making under uncertainty: A comparison of four models. *Organizational Behavior and Human Decision Processes*, *74*(2), 118–144.

Barron, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making*, *16*(3), 215–233.

Barron, G., Leider, S., & Stack, J. (2008). The effect of safe experience on a warnings' impact: Sex, drugs, and rock-n-roll. *Organizational Behavior and Human Decision Processes*, *106*(2), 125–142.

Barron, G., & Yechiam, E. (2009). The coexistence of overestimation and underweighting of rare events and the contingent recency effect. *Judgment and Decision Making*, *4*(6), 447–460.

Becker, C. (1968). Punishment: An Economic Approach. *Journal of Political Economy*, *76*(2), 169–217.

Bereby-Meyer, Y., & Erev, I. (1998). On learning to become a successful loser: a comparison of alternative abstractions of learning processes in the loss domain. *Journal of Mathematical Psychology*, *42*(2), 266–286.

Camerer, C., & Weber, M. (1992). Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of Risk and Uncertainty*, *5*(4), 325–370. http://doi.org/10.1007/BF00122575

Camilleri, A. R., & Newell, B. R. (2009). The role of representation in experience-based choice. *Judgment and Decision Making*, *4*(7), 518–529.

Denrell, J., & March, J. G. (2001). Adaptation as information restriction: The hot stove effect. *Organization Science*, *12*(5), 523–538.

Edwards, W. (1961). Probability learning in 1000 trials. *Journal of Experimental Psychology*, *62*(4), 385.

Ellsberg, D. (1961). Risk, Ambiguity, and the Savage Axioms. *The Quarterly Journal of Economics*, *75*(4), 643–669.

Erev, I., & Barron, G. (2005). On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological Review*, *112*(4), 912.

Erev, I., Ert, E., Plonsky, O., Cohen, D., & Cohen, O. (2017). From anomalies to forecasts: a choice prediction competition for decisions under risk and ambiguity. *Psychological Review*, *124*(4)369–409.

Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S. M., Hau, R., Hertwig, R., Stewart, T., West, R., & Lebiere, C. (2010). A choice prediction competition: choices from experience and from description. *Journal of Behavioral Decision Making*, *23*(1), 15–47.

Erev, I., Ert, E., & Yechiam, E. (2008). Loss aversion, diminishing sensitivity, and the effect of experience on repeated decisions. *Journal of Behavioral Decision Making*, *21*(5), 575–597. http://doi.org/10.1002/bdm.602

Erev, I., & Rodensky, D. (2004). Gentle enforcement of safety rules. *A Final Report of a Research Supported by the Committee for Accident Prevention in the Israeli Ministry of Industry & Commerce.* Technion, Haifa, Israel (in Hebrew).

Ert, E., & Erev, I. (2007). Replicated alternatives and the role of confusion, chasing, and regret in decisions from experience. *Journal of Behavioral Decision Making*, *20*(3), 305–322. http://doi.org/10.1002/bdm.556

Ert, E., Raz, O., & Heiman, A. (2016). (Poor) seeing is believing: When direct experience impairs product promotion. *International Journal of Research in Marketing*, *33*(4), 881–895.

Ert, E., & Trautmann, S. T. (2014). Sampling experience reverses preferences for ambiguity. *Journal of Risk and Uncertainty*, *49*(1), 31–42. http://doi.org/10.1007/s11166-014-9197-9

Estes, W. K., & Suppes, P. (1959). *Foundations of statistical learning theory. II. The stimulus sampling model.* Stanford University, Applied Mathematics and Statistics Laboratory, Behavioral Sciences Division. Retrieved from http://suppes-corpus.stanford.edu/techreports/IMSSS_26.pdf

Fox, C. R., & Tversky, A. (1998). A belief-based account of decision under uncertainty. *Management Science*, *44*(7), 879–895. http://doi.org/10.1287/mnsc.44.7.879

Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology*, *65*(1), 45–55.

Fujikawa, T. (2009). On the relative importance of the hot stove effect and the tendency to rely on small samples. *Judgment and Decision Making*, *4*(5), 429–435.

Grant, D. A., Hake, H. W., & Hornseth, J. P. (1951). Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *Journal of Experimental Psychology*, *42*(1), 1–5.

Hadar, L., & Fox, C. R. (2009). Information asymmetry in decision from description versus decision from experience. *Judgment and Decision Making*, *4*(4), 317–325.

Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description-experience gap in risky choice: the role of sample size and experienced probabilities. *Journal of Behavioral Decision Making*, *21*(5), 493–518. http://doi.org/10.1002/bdm.598

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*(8), 534–539. http://doi.org/10.1111/j.0956-7976.2004.00715.x

Hertwig, R., & Erev, I. (2009). The description-experience gap in risky choice. *Trends in Cognitive Sciences*, *13*(12), 517–523. http://doi.org/16/j.tics.2009.09.004

Hills, T. T., Noguchi, T., & Gibbert, M. (2013). Information overload or search-amplified risk? Set size and order effects on decisions from experience. *Psychonomic Bulletin & Review*, *20*(5), 1023–1031.

Jessup, R. K., Bishara, A. J., & Busemeyer, J. R. (2008). Feedback produces divergence from prospect theory in descriptive choice. *Psychological Science*, *19*(10), 1015–1022. http://doi.org/10.1111/j.1467-9280.2008.02193.x

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, *47*(2), 263–291.

Lejarraga, T., & Gonzalez, C. (2011). Effects of feedback and complexity on repeated decisions from description. *Organizational Behavior and Human Decision Processes, 117,* 286–295.

Marchiori, D., Di Guida, S., & Erev, I. (2015). Noisy retrieval models of over-and undersensitivity to rare events. *Decision*, *2*(2), 82–106.

Munichor, N., Erev, I., & Lotem, A. (2006). Risk attitude in small timesaving decisions. *Journal of Experimental Psychology: Applied*, *12*(3), 129–141.

Nevo, I., & Erev, I. (2012). On surprise, change, and the effect of recent outcomes. *Frontiers in Psychology*, *3*. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3283116/

Pauly, M. V. (1968). The Economics of Moral Hazard: Comment. *The American Economic Review*, *58*(3), 531–537.

Phillips, L. D., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, *72*(3), 346.

Pizam, A., & Fleischer, A. (2002). Severity versus frequency of acts of terrorism: Which has a larger impact on tourism demand? *Journal of Travel Research*, *40*(3), 337–339.

Schurr, A., Rodensky, D., & Erev, I. (2014). The effect of unpleasant experiences on evaluation and behavior. *Journal of Economic Behavior & Organization*, *106*, 1–9.

Shafir, S., Reich, T., Tsur, E., Erev, I., & Lotem, A. (2008). Perceptual accuracy and conflicting effects of certainty on risk-taking behaviour. *Nature*, *453*(7197), 917–920.

Siegel, S. (1961). decision making and learning under varying conditions of reinforcement. *Annals of the New York Academy of Sciences*, *89*(5), 766–783.

Simmons, K. M., & Sutter, D. (2009). False alarms, tornado warnings, and tornado casualties. *Weather, Climate, and Society*, *1*(1), 38–53.

Ungemach, C., Chater, N., & Stewart, N. (2009). Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)? *Psychological Science*, *20*(4), 473–479. http://doi.org/10.1111/j.1467-9280.2009.02319.x

Vulkan, N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys*, *14*(1), 101–118.

Weber, E. U., Shafir, S., & Blais, A. R. (2004). Predicting risk sensitivity in humans and lower animals: risk as variance or coefficient of variation. *Psychological Review*, *111*(2), 430–445.

Weimann, G. (1993). *The Theater of Terror: Mass Media and International Terrorism*. New York: Longman.

Yechiam, E., Barron, G., & Erev, I. (2005). The role of personal experience in contributing to different patterns of response to rare terrorist attacks. *Journal of Conflict Resolution*, *49*(3), 430–439. http://doi.org/10.1177/0022002704270847

Yechiam, E., & Busemeyer, J. R. (2006). The effect of foregone payoffs on underweighting small probability events. *Journal of Behavioral Decision Making*, *19*(1), 1–16.

Yechiam, E., Druyan, M., & Ert, E. (2008). Observing others' behavior and risk taking in decisions from experience. *Judgment and Decision Making*, *3*(7), 493–500.

Yechiam, E., Erev, I., & Barron, G. (2006). The effect of experience on using a safety device. *Safety Science*, *44*(6), 515–522.

Zohar, D. (1980). Safety climate in industrial organizations: theoretical and applied implications. *Journal of Applied Psychology*, *65*(1), 96–102.

Zohar, D., & Erev, I. (2007). On the difficulty of promoting workers' safety behavior: overcoming the underweighting of routine risks. *International Journal of Risk Assessment and Management*, 7(2), 122–136.

Zweifel, P., & Manning, W. G. (2000). Chapter 8 Moral hazard and consumer incentives in health care. In Anthony J. Culyer and Joseph P. Newhouse (Eds.), *Handbook of Health Economics* (Vol. Volume 1, Part A, pp. 409–459). Elsevier. Retrieved from http://www.sciencedirect.com/science/article/pii/S1574006400801675

8 Social amplification of risk and extreme events

Roger E. Kasperson

A perplexing problem in risk analysis is why some relatively minor hazards or hazard events (as assessed by technical experts) elicit strong public concerns and why other extreme hazards are tolerated and generate little public reaction. (Here, I am defining "hazards" as threats to people and what they value. "Hazard events" include not only actual accidents or consequential events, but also reports that characterize the hazards.) Such effects are typically the result of "social amplification"—changes in risk perception and response based on psychological, social, institutional, and cultural processes (Fischhoff 1995). Social amplification is most likely to flourish when the hazards are either serious or extreme and the situation is fraught with uncertainties.

This article describes an approach, the social amplification of risk framework, for understanding and accounting for public attitudes and responses toward risk (Kasperson and Kasperson 1996). The framework set forth links the technical assessment of risk with psychological, sociological, and cultural perspectives of risk perception and risk-related behavior (Kasperson et al. 1988; Pidgeon et al. 2003; Teversky and Kahneman 1974). The main thesis of the framework is that hazards interact with these perspectives in ways that may amplify or attenuate public responses. In this chapter, I focus on social amplification of extreme hazard events. Many articles have appeared in the risk field treating this subject (see Appendix 8.1).

Risk amplification typically occurs at two stages in a risk scenario: in the transfer of information about risk and in social response mechanisms. Signals about risk are both transmitted and processed by individuals and social entities, called "amplification stations." The individual might be a scientist, for example, who communicates the risk assessment; a social entity might be the news media, a cultural group, or an interpersonal network (Hill 2001). The perceived amplified risk leads to behavioral responses that result in secondary impacts or "ripples."

Social amplification may qualitatively and quantitatively increase not only the perception of risk but also the risk itself and its consequences (Funtowicz and Ravetz 1990). For this reason, social amplification of risk must be included in analyses of public and regulatory reactions to risk events (but see Rayner 1988).

The key amplification stages are listed below:

- filtering signals (only a fraction of all incoming information is actually processed);
- decoding and reframing the signals;
- processing risk information (e.g. drawing inferences);
- attaching social values to information as a basis for drawing implications for management and policy;

- interacting with cultural and peer groups to interpret and assess the validity of signals;
- formulating behavioral intentions about whether to tolerate a risk or take action against the risk or risk manager;
- engaging in group or individual actions to accept, ignore, tolerate, or change the risk.

An extreme event—the accident at the Three Mile Island (TMI) nuclear reactor in 1979—demonstrated dramatically that factors besides injury, death, and property damage can impose serious costs and social repercussions. While no one likely died from the release of radioactivity at TMI, few accidents in U.S. history have wrought such costly societal impacts. The TMI accident, and the more recent event at Fukushima in Japan, devastated the utilities that owned and operated the plants and imposed enormous costs—in the form of stricter regulations, reduced operation of reactors worldwide, greater public opposition to nuclear power, and a less viable role for one of the major long-term energy sources—at a time when nuclear power could play an important role in climate change. These events may also have increased public concerns about other complex technologies, such as chemical manufacturing and genetic engineering.

The point is that traditional cost-benefit and risk analyses neglect these higher-order impacts and thus can greatly underestimate the variety of adverse effects attendant on certain hazard events (and thereby underestimate the overall risk from the event) (Dietz and Stern 1996). In this sense, social amplification provides a corrective mechanism by which society acts to bring the technical assessment of risk more in line with a fuller determination of risk (Funtowicz and Ravetz 1990). At the other end of the spectrum, the relatively low levels of interest by the public in the risks presented by such well-documented and significant hazards such as fracking, smoking, or driving while texting serve as examples of potential social attenuation of risk. Whereas attenuation of risk is indispensable in that it allows individuals to cope with the multitude of hazards and hazard events encountered daily (Kahneman 2011), it also may lead to potentially serious adverse consequences from underestimation and under-response (Slovic 2010). Thus, both social amplification and attenuation, through serious disjunctures between expert and public assessments of risk and varying responses among different publics, confound conventional risk analysis. In some cases, the societal context may, through its effects on the risk assessor, alter the focus and scope of the risk assessment (National Research Council 2009).

The social amplification of risk approach can, in principle, provide a theoretical base for a more comprehensive and powerful analysis of risk and risk management in modern societies. At this point, we propose a conceptual framework that may serve to guide ongoing efforts in develop, test, and apply such an approach to extreme events.

The social amplification framework

Social amplification of risk denotes the phenomenon by which information processes, institutional structures, social group behavior, and individual responses shape the social experience of risk, thereby contributing to risk consequences (see Figure 8.1).

Thus there is no such thing as 'true' (absolute) and 'distorted' (socially determined) risk. Rather the information system and public responses all are essential elements in determining the nature and magnitude of risk (Dietz and Stern 2008).

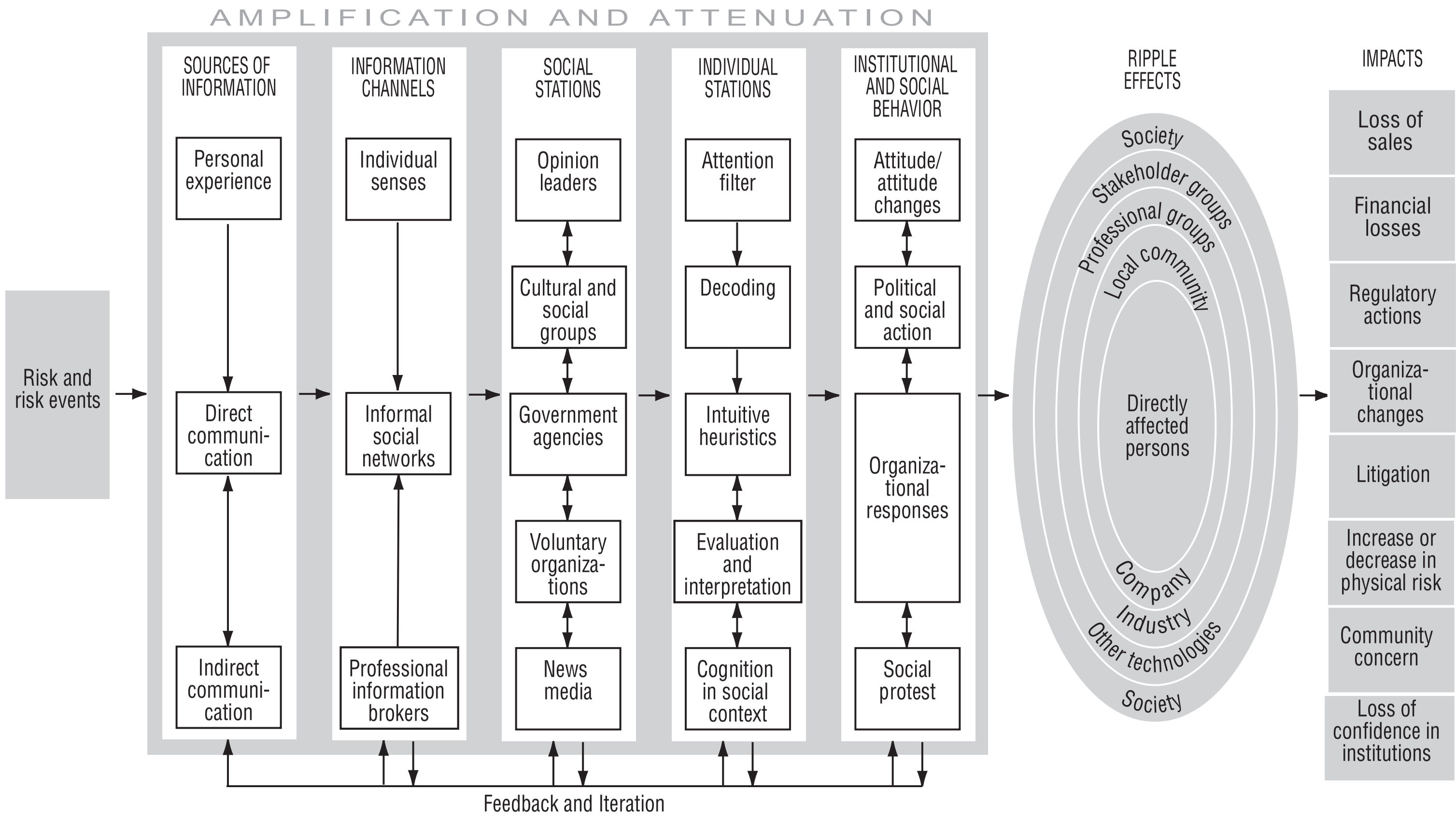


Figure 8.1 Detailed conceptual framework of social amplification of risk.

Like a stereo receiver, the information system may amplify hazard events in two ways:

- by intensifying or weakening signals that are part of the information that individuals and social groups receive about the hazard;
- by filtering the multitude of signals with respect to the attributes of the hazard and their importance (Renn 1991; Renn et al. 1993).

Signals arise through direct personal experience with a risk object or through the receipt of information about the hazard from the mass or social media or groups. These signals are processed by social, as well as individual, amplification 'stations' which include:

- the scientist who conducts and communicates the technical assessment of risk;
- risk management institutions;
- social media;
- the news media;
- activist social organizations;
- opinion leaders within social groups;
- personal networks of peer groups;
- public agencies.

Social amplification stations generate and transmit information via communications channels (mass media, social media, telephones, and direct conversations). In addition, each recipient also engages in amplification and attenuation processes, thereby acting as an amplification station for risk-related presentations. Ultimately, we seek to know why specific hazards and hazard events undergo more or less amplification or attenuation.

Ripple effects

Imagining ripples in a pond is a good way to think about how impacts associated with the social amplification of risk spread outward or stay inward (see the right-hand side of Figure 8.1 above), affecting the physical environment, social apathy, and stigmatization of an environment of a risk manager, and producing

- losses in local business sales, lower residential property values, and lower levels of economic activity;
- political and social pressure (e.g., political demands, changes in the political climate and culture);
- changes in the nature of the risk (e.g., feedback mechanisms that heighten or lower the risk);
- changes in training, education, or required qualifications for operations and emergency response personnel; and even
- social disorder (e.g., protests, riots, sabotage, terrorism);
- changes in risk monitoring and regulation;
- higher liability and insurance costs;
- repercussions on other technologies (e.g., lower levels of public acceptance) and on social institutions (e.g., erosion of public trust), as when the 1989 explosion at the chemical plant in Bhopal, India, raised concerns about the possible failure of "fail-safe" systems at nuclear power plants (Kasperson et al. 1992).

Once secondary impacts are perceived by social groups and individuals, they may lead to another stage, or to the next, such as an institutional level (a company or a government regulatory agency).

In the wake of the Deepwater Horizon explosion in April 2010, for example, the effects spread from the drilling rig to the rest of the Gulf of Mexico, including the wetlands and beaches in all of the adjacent states, and then to politicians and petroleum industry representatives who were compelled to reconsider their plans to expand offshore drilling.

The concept of rippling impacts suggests processes that can extend (in risk amplification) or constrain (in risk attenuation) temporal, sectoral, and geographical impacts. It also illustrates that each order of impact, or ripple, may not only have social and political effects but also trigger (in risk amplification) or hinder (in risk attenuation) managerial interventions to reduce risk. Secondary effects include market impacts (e.g., consumer avoidance of a product or related products), demands for regulatory constraints, litigation, community opposition, loss of credibility and trust, stigmatization of a facility or community, and investor flight.

Each order of impact may also trigger (in risk amplification) or hinder (in risk attenuation) positive changes for risk reduction. Positive changes were apparent, for example, in the wake of the Fukushima nuclear accident in Japan when Germany restructured its energy system and the United States (among other countries) launched a major (and costly) review of all its nuclear plants.

As yet, limited attention has addressed the role of organizations and institutions in the social processing of risk (Dietz and Stern 1996). Pidgeon (1997) has suggested that linking risk amplification to the considerable empirical base of knowledge concerning organizational processes intended to prevent large-scale failures and disasters would be an important extension of the framework. Many contemporary hazards originate in social technical systems (see Turner 1978; Short and Clarke 1992) rather than natural phenomena so that risk management and internal regulatory processes governing the behavior of institutions in identifying, diagnosing, prioritizing, and responding to hazards are key parts of the broader amplification process. Large organizations increasingly set the context and terms of debate for society's consideration of risk. Understanding amplification dynamics, then, requires insight into how risk-related decisions relate to organizational self-interest, messy inter- and intra-organizational relationships, economically related rationalizations, and "rule of thumb" considerations that often conflict with the view of risk analysis as a scientific enterprise (Short 1992; Short and Clark 1992). Since major accidents are often preceded by smaller incidents and risk warnings, how signals of incubating hazards are processed within institutions and communicated to others outside the institution does much to structure society's experience with technological and industrial hazards.

Noting the relative void of work on organizational risk processing, Freudenburg (1992) has examined characteristics of organizations that attenuate risk signals and ultimately increase the hazards posed by technological systems and notably extreme events. These include such attributes as the lack of organizational commitment to the risk management function, the bureaucratic attenuation of information flow within the organization (and particularly in a "bad news" context), specialized divisions of labor that create "corporate gaps" in responsibility, amplified risk-taking by workers, the atrophy of organizational vigilance to risk as a result of a myriad of factors (e.g. boredom, routinization), and imbalances and mismatches in institutional resources. Such organizational attenuation of risk systematically amplifies the health and environmental hazards that the organization is entrusted to anticipate and to control.

Other studies of organizational handling of risk provide further considerations. In an analysis of safety management at Volvo, Svenson (1988) found that patterns of access of various parties to reliable information about hazards, understanding of the relation between changes in the product and changes in safety levels, attributes of the information network that informs various interested parties about changes in safety (either positive or negative), and the presence of organizational structures for negotiating about risk between the producer and other social institutions were all important. In her analysis of the *Challenger* accident in the United States, Vaughan (1992) also found communication and information issues to be critical but argued that structural factors, such as pressures from a competitive environment, resource scarcity in the organization, increased the vulnerability of individuals and systems at risk.

The organizational amplification of risk

Several theoretical perspectives on organizational processing of risk can be drawn upon within the amplification/attenuation framework. When considering individual responses to hazards within organizations, the idea of psychological denial of threatening information has had a particularly long history. Perhaps surprisingly, however, this topic has only rarely been investigated in contemporary psychological risk research, where the current paradigm for understanding responses to hazard and environmental threat, based upon the two traditions of cognitive psychology and decision making research, has long undervalued the role of motivational variables in behavior and choice (with the notable exceptions of Slovic 1987 and Fischhoff 1995). Recent work has now begun to explore the relationship between affective variables and risk perceptions (Slovic 2000; 2010).

Evidence of a range of broad social and organizational preconditions to large-scale accidents is available in the work of Turner (1978). As a result of a detailed analysis of 84 major accidents in the United Kingdom, Turner concluded that such events rarely come about for any single reason. Rather, it is typical to find a number of undesirable events that accumulate, unnoticed or not fully understood, often over a considerable number of years, which he defines as the *disaster incubation period*. Preventive action to remove one or more of the dangerous conditions or a *trigger event,* which might be a final critical error or a slightly abnormal operating condition, brings this period to an end. Turner focuses in particular upon the information difficulties, which are typically associated with the attempts of individuals and organizations to deal with uncertain and ill-structured safety problems, during the hazard-incubation period (see also Johnson and Slovic 1995).

Despite these valuable explorations, our knowledge of risk amplification and attenuation in different types of institutions remains thin and eclectic. Systematic application of the amplification framework in a comparative study of organizations, issues, and institutions (see e.g. the work comparing regulation of radon, chemicals, and BSE by Rothstein 2003) might well yield highly useful results, particularly demonstrating how signals are denied, deemphasized, or misinterpreted.

Managing risk amplification and attenuation

Responding to amplified and attenuated risk requires attention to the social, institutional, and cultural contexts in which risk signals are decoded and management options assessed. Four major pathways exist for framing decision problems and priorities:

1 *Heuristics and values*. Individuals cannot deal with the full complexity of risk and the multitude of hazards involved in daily life. Thus people use simplifying mechanisms to evaluate risk and to shape responses. These processes, while permitting individuals to cope with a risky world, can sometimes introduce biases that cause distortions and errors (Kahneman 2011). The application of individual and group values will often determine which hazards are deemed important or minor and what actions, if any, should be taken.

2 *Social group relationships*. Risk issues enter into the political agenda of social and political groups. The nature of these groups will influence member responses and the types of rationality brought to risk issues. To the extent that risk becomes a central issue in a political campaign or in a conflict among social groups, it will be vigorously brought to more general public attention, often coupled with ideological interpretations of technology or the risk management process (Johnson and Covello 1987). A polarization of views and an escalation of rhetoric by partisans typically occur, and new recruits are drawn into the conflict. These social alignments tend to become anchors for subsequent interpretations of risk management and may become quite firm in the face of conflicting information.

3 *Signal value*. An important concept that has emerged from research on risk perception is that the seriousness and higher-order impacts of a hazard event are determined, in part, by what that event signals or portends (Slovic 1987). The informativeness or 'signal value' of an event, whether extreme or more common, appears to be systematically related to the characteristics of the event and the hazard it reflects. High-signal events may suggest that a new hazard has appeared or that the risk is different and more serious than previously understood. Thus an accident that takes many lives may produce relatively little social disturbance (beyond that experienced by the victims' families and friends) if it occurs as part of a familiar and well-understood system (such as a train wreck). An accident in an unfamiliar system, such as a nuclear reactor or a recombinant-DNA laboratory, however, may elicit great public concern if it is interpreted to mean that the hazard is not well understood, not controllable, or not competently managed, thus implying that further (and possibly worse) mishaps are likely. In sum, signals about a hazard event initiate a process whereby the significance of the event is publicly examined. If found to be ominous, these implications are likely to trigger higher-order social and economic impacts.

4 *Stigmatization*. Stigma refers to the negative imagery associated with undesirable social groups or individuals (Goffman 1963). But environments with heavy pollution, waste accumulation, or hazardous technology may also come to be associated with negative images. Love Canal, Fukushima, the Valley of the Thousand Drums, Times Beach, and the Nevada Test Site evoke vivid images of threat and pollution. Since the typical response to stigmatized persons or environments is avoidance, it is reasonable to assume that risk-induced stigma may have significant social and policy consequences (Flynn et al. 2001). Figure 8.2 provides an indication of how amplification can be applied to analyze stigma (see also Kasperson et al. 2001).

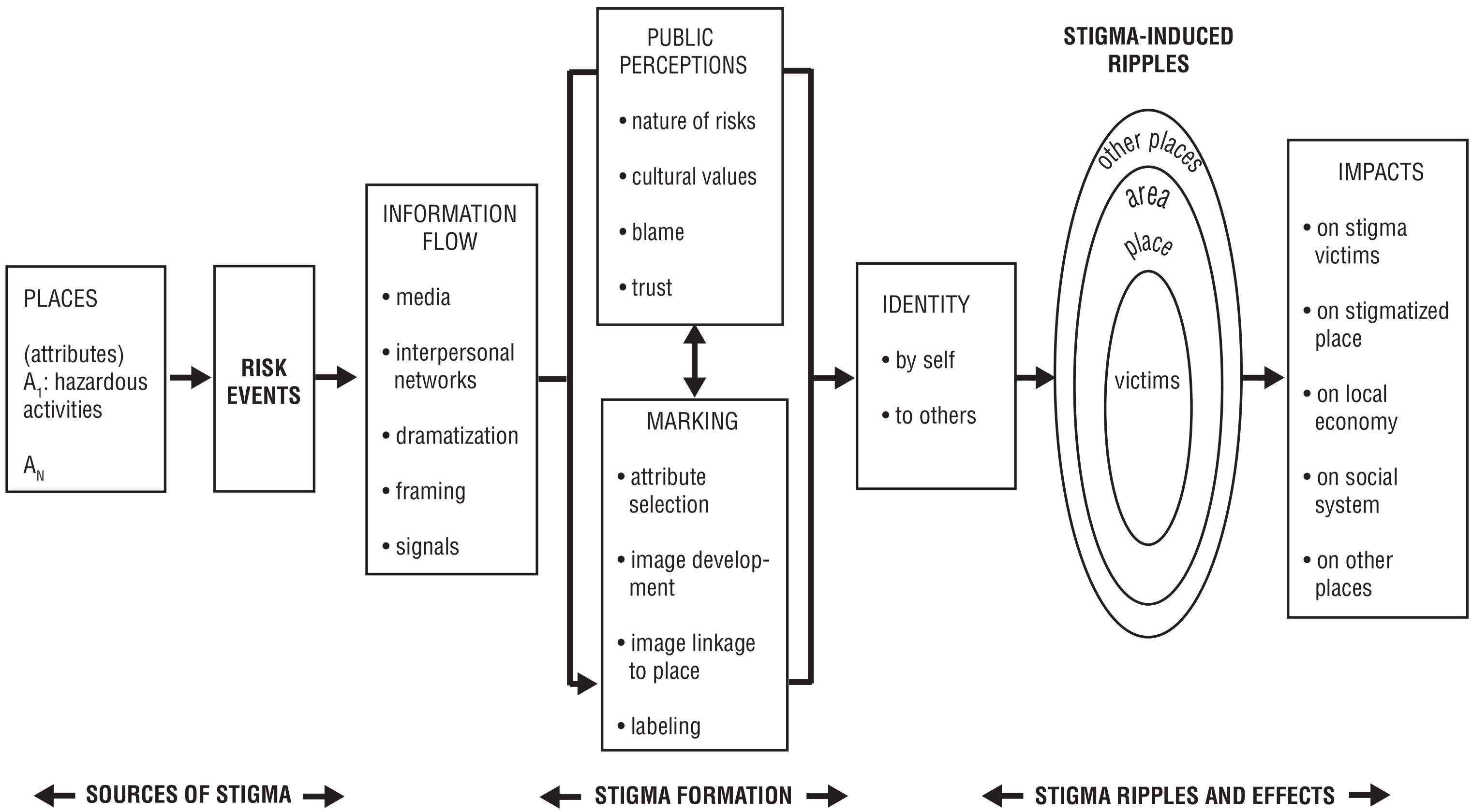


Figure 8.2 Risk amplification and stigmatization (Kasperson et al. 2001, p. 20).

In addition to these four pathways or mechanisms, *positive feedback to the physical hazard itself* can occur due to social processes. If a transportation accident with hazardous materials were to occur close to a waste disposal site, for example, protests and attempted blockage of the transportation route could result. Such actions could themselves become initiating or co-accident events, thereby increasing the probabilities of future accidents or enlarging the consequences should an accident occur. Especially where strong public concern exists over an environmental threat, or technology, a wide variety of mechanisms is present by which health and safety hazards may be enlarged through social processes.

Observations on three extreme events

Extreme events pose perplexing issues for decision makers. Inevitably, extreme events are highly uncertain and often unanticipated. So frequently, preparations have been minimal. And yet society gleans major messages from such experience. Social processes interact with biophysical and medical hazards to change perception of what the hazards are and how they may best be managed. Inevitably, uncertainties as to what the future holds abound. In the three cases that follow, no attempt is made to provide a full analysis of the amplification and attenuation that such events engender. Rather, observations are made as to how social amplification and attenuation processes have shaped society's experience of such events and entered into management and decision making. Many candidates exist for analysis, as this book makes clear, but here three extreme events—the BP Gulf oil spill in the U.S., the Fukushima nuclear disaster in Japan, and the 2008 Wenchan earthquake in China are examined for major insights.

The BP oil spill

On April 20, 2010, the Deepwater Horizon offshore drilling unit exploded and resulted in a massive oil spill in the Gulf of Mexico. Some 11 workers died in the event. A host of coastal government and economy sectors brought criminal charges against the petroleum giant. Eventually these suits were settled by BP agreeing to pay $2.5 billion to jump start affected ecological restoration needs in the five Gulf States. Another $350 million was allocated to the U.S. National Academy of Sciences to create a three-decades long research effort on potential ecological damage, $2.4 billion to the National Fish and Wildlife Foundation, $500 million to the Gulf of Mexico Research Initiative and another $100 million to the American Wetlands Conservation Fund (Malakoff 2012). In short, major agents of expertise and amplification analysis were brought into the long-term research initiative on risk.

Social response to what may well be the worst environmental disaster in the U.S. has been commensurate with the importance of the event. Extensive media coverage has occurred and thousands of court suits for damages have been brought against BP. Yet the debate has been chiefly one of finger-pointing: was BP primarily to blame or was it Trans Ocean, the driller? The reality, of course, is that Louisiana, the hardest hit state, has a population on the one hand often remote from coastal locations and on the other hand heavily reliant upon Gulf coastal ecosystems and resources. Yet the ongoing debate is largely about whether local claims for damages are valid or fraudulent. The U.S. Congress, ever protective of the petroleum industry, has primarily focused on the allocation of blame. Meanwhile, the larger issues—petroleum and climate change, offshore drilling, and national energy consumption—have largely been attenuated (but see Freudenburg et al. 2009). But overall, neither risk amplification nor attenuation appears to have been decisive in this case.

The Wenchan earthquake

In 2008, an earthquake of 8.0 magnitude struck Sichuan, China, resulting in 70,000 lives lost and 1.5 trillion RMB in direct economic losses (Yonge and Booth 2011). While Chinese seismologists were active and often impressive in earthquake prediction, the Wenchan earthquake came as a surprise to risk managers and experts alike. Occurring in a densely populated and mountainous area, a high loss of life, much of it from hazard chain effects (landslides, rock falls, and mud floods), were unanticipated results. Meanwhile, thousands of buildings collapsed, causing further loss of life and impeding response.

The lack of prior preparation particularly in the needed strengthening of buildings (especially in rural zones), obviously played a major role in shaping the disaster toll. But social amplification played a major role in disaster response as well. The rapid coverage of the event in broadcast reports and extensive film footage stimulated an unprecedented flow of risk signals and added new dimensions to societal response. Meanwhile, disaster relief and teams of volunteers from China and abroad provided urgently needed financial, medical, and psychological resources.

During the disaster, and shortly thereafter, ad hoc responses filled some of the gap apparent in preplanning of information flow during natural disasters. The Chinese government played its role. In less than 10 minutes after the earthquake struck, the Chinese Earthquake Administration linked with the Xinhua News Agency to provide ongoing effects and responses to communities at risk. Central Television immediately began to provide live coverage of relief efforts. The *People's Daily* and other newspapers moved quickly to provide ongoing coverage. Some 150 reporters from foreign media from more than 20 countries arrived in the disaster area to provide worldwide coverage of the event and its aftermath. And, for the first time, social media played an important role. The Internet, for example, was extensively used to provide needed information to assist rescue and recovery efforts. All this may well have enhanced greater public trust in the media and government (Yong and Booth 2011).

But perhaps most important was the social learning that resulted from the event. In the months following the Wenchan disaster, the Chinese government adopted a "peoples first open-information" response to natural disasters, as well as adoption of stricter building codes and other disaster preparatory and mitigation measures. Already the government adopted emergency response provisions and communication upgrades following the devastating Tangshan earthquake of 1976. Despite this, the Wenchan earthquake drove home the need for greater measures to reduce vulnerabilities (especially in buildings) and to improve emergency communications during and following disaster events. While social learning from earthquake experience clearly occurred and has been translated into policy, risk attenuation has also occurred in risk preparatory measures and failures in emergency communication. Future events will provide enhanced understanding of how social processes can contribute to risk reduction rather than a reliance on removing the many uncertainties in earthquake risk prediction.

The Fukushima Daiichi disaster

On March 11, 2011 a 9.0 magnitude earthquake struck off the east coast of Japan. The earthquake triggered a huge tsunami, 46 minutes after the earthquake whose 45-foot waves overwhelmed existing seawalls, including those at the Fukushima Daiichi nuclear plant. The immediate results, prior to the chain of disaster that eventually developed (as with the Wenchan earthquake), involved some 15,000 people killed, 170,000 persons evacuated, and a loss of power at the Fukushima Daiichi plant.

The event soon cascaded into a series of other events at the nuclear plant. Plant operators faced a situation of no power or reactor control, a loss of cooling water, and widespread communication failures inside the plant and with surrounding communities (Beyea et al. 2013). Augmenting these problems was the absence of any disaster management plan, so that responses were ad hoc and often reactive. Dominant throughout, contrary to extensive social science research (Mileti 1999) was the technical group concern at Fukushima that the public would panic and engage in antisocial behavior. As a result, information needed by publics at risk was not forthcoming, but held back by the concern of possible panic (and thereby contributing to risk amplification).

A large and largely unplanned evacuation did occur, affecting communities within a 30-kilometer radius of the plant (Craft 2011). In all some 170,000 persons were evacuated, and some 600 deaths resulting from the evacuation. Evacuees were housed in very primitive conditions and usually not allowed to return to their homes. The grieving for a lost home and community was one of the major results of risk management and risk amplification.

The extensive growth of public concern in Japan as a result of the extensive amplification of nuclear risk not only resulted in the shutdown of all of Japan's extensive network of nuclear reactors but also produced widespread stigma effects. Japan has witnessed a widespread decline in worldwide consumer demand for its agricultural products, due to radioactive contamination fears. In Japan, restrictions were also placed on food and water consumption (Farber 2012). This has even resulted in testing of marine and land agricultural products due to the leaks and releases of radiation into the ocean on the west coast of the U.S. At the same time, a loss of trust among the Japanese public in nuclear power and its management institutions has been apparent, due in no small part to inadequacies in information flow and a lack of attention to public concerns and the continuing discovery of new hazard elements (such as "hot spots" in the ocean floor near Fukushima).

The ripple effects associated with the worldwide amplification of the accident risks have become apparent internationally. Germany announced its intent to phase out its national system of 50 nuclear plants by 2022. The United States undertook a major review of all its nuclear plants to determine if their safety systems could deal with the risk challenges apparent in the Fukushima Daiichi event. Meanwhile, the ripples have extended to patterns of public support and opposition to nuclear power at a time of climate change considerations (Butler et al. 2011). Worldwide plant operators and regulators are reviewing their safety measures in light of the amplified public concerns.

Conclusion

It is abundantly apparent that societal experience of extreme events is a product not only of the public health and ecological implications but of how events connect to ongoing social processes. Everywhere the cry is out for integrated risk and environmental assessment. Anything less frequently misses how extreme and even minor events are experienced and the entrance of such events into risk management programs and priorities. In short, such events must be analyzed in terms of their risk signal attributes and the ripple effects which result. The people-related issues in risk experience have too long been neglected or only partially understood. Extreme events have a large potential to reshape this situation.

References

Beyea, J., E. Lyman and F.N. von Hippel. 2013. Accounting for long-term doses in "worldwide health effects of the Fukushima Daiichi nuclear accident." *Energy and Environmental Science* 6 (3): 1042–1045.

Butler, C., K.A. Parkhill and H.F. Pidgeon. 2011. Nuclear power after Japan: The social dimensions. *Environment* 53 (6): 3–14.

Craft, L. 2011. Japan's nuclear refugees. *National Geographic* (December): 91–109.

Dietz, T. and P. Stern, eds. 1996. *Understanding Risk*. Washington, DC: National Academy Press.

Dietz, T. and P. Stern, eds. 2008. *Public Participation in Environmental Assessment and Decision Making*. Washington, DC: National Academies Press.

Farber, D.A. 2012. Introduction: Legal scholarship, the disaster cycle, and the Fukushima accident. *Duke Environmental Law and Policy Review* 23 (1): 1–21.

Fischhoff, B. 1995. Risk perception and risk communication unplugged: Twenty years of process. *Risk Analysis* 15 (2): 137–145.

Flynn, J., P. Slovic and H. Kunreuther, eds. 2001. *Risk, Media and Stigma: Understanding Public Challenges to Modern Society and Technology*. London: Earthscan.

Freudenburg, W.R. 1992. Nothing recedes like success? Risk analysis and the organizational amplification of risks. *Risk: Issues in Health and Safety* 3: 1–35.

Freudenburg, W.R., R. Gramling, S. Laska, and K. Erikson. 2009. *Catastrophe in the Making*. Washington, DC: Island Press, 2009.

Funtowicz, S.O. and J. Ravetz. 1990. *Global Environmental Issues and the Emergence of Second Order Science*. London: Council for Science and Society.

Goffman, E. 1963. *Stigma: Notes on the Management of Spoiled Identity*. New York: Simon and Schuster.

Hill, A. 2001. Media risks: The social amplification of risk and the media debate. *Journal of Risk Research* 4: 209–226.

Johnson, B.B. and V.S. Covello, eds. 1987. *The Social and Cultural Construction of Risk*. Dordrecht, the Netherlands: Reidel.

Johnson, B.B. and P. Slovic. 1995. Presenting uncertainty in health risk assessment: Initial studies of its effects on risk perception and trust. *Risk Analysis* 15 (4): 485–494.

Kahneman, D. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus, and Giroux.

Kasperson, R.E., D. Golding and S. Tuler. 1992. Social distrust as a factor in siting hazardous facilities and communicating risks. *Journal of Social Issues* 48 (4): 161–187.

Kasperson, R.E., N. Jhaveri and J.X. Kasperson. 2001. Stigma, places, and the social amplification of risk: Toward a framework of analysis. In *Risk, Media and Stigma: Understanding Public Challenges to Modern Science and Technology,* edited by J. Flynn, P. Slovic, and H. Kunreuther, 9–27. London: Earthscan.

Kasperson, J.X. and R.E. Kasperson. 1996. The social amplification and attenuation of risk. *The Annals of the American Academy of Political and Social Science* 545: 95–105.

Kasperson, R.E., O. Renn, P. Slovic, H.S. Brown, J. Emel, R. Goble, J.X. Kasperson and S.J. Ratick. 1988. The social amplification of risk: A conceptual framework. *Risk Analysis* 8 (2): 178–187.

Malakoff, D. 2012. BP criminal case generates record payout for science and restoration. *Science* 338 (6111): 1137. DOI: 10.1126/science.338.6111.1137

Mileti, D. 1999. *Disasters by Design: A Reassessment of Natural Hazards in the United States*. Washington, DC: John Henry Press.

National Research Council (NRC). 2009. *Science and decisions: Advancing Risk Assessment*. Washington, DC: National Academies Press.

Pidgeon, N. 1997. *Risk communication and the social amplification of risk: Phase 1*, Scoping Study. Report to the UK Health and Safety Executive. London: HSE Books.

Pidgeon, N., R.E. Kasperson and P. Slovic, eds. 2003. *The Social Amplification of Risk*. Cambridge, UK: Cambridge University Press.

Rayner, S. 1988. Muddling through metaphors to maturity: A commentary on Kasperson et al. 'The Social Amplification of Risk.' *Risk Analysis* 8 (2): 201–204.

Renn, O. 1991. Risk communication and the social amplification of risk. In *Communicating Risks to the Public International Perspectives,* edited by R.E. Kasperson and P.J.M. Stallen, 287–324. Dordrecht, the Netherlands: Kluwer Academic Press.

Renn, O., W.J. Burns, J.X. Kasperson, R.E. Kasperson and P. Slovic. 1993. The social amplification of risk: Theoretical foundations and empirical applications. *Journal of Social Issues* 48 (4): 137–160.

Rothstein, H. 2003. Neglected risk regulation: The institutional attenuation phenomenon. *Health, Risk and Society* 5: 85–103.

Short, J.F., Jr. 1992. Defining, explaining, and managing risk. In *Organizations, Uncertainties, and Risk*, edited by J.F. Short, Jr. and L. Clarke, 3–23. Boulder, CO: Westview Press.

Short, J.F., Jr. and L. Clarke, eds. 1992. *Organizations, Uncertainties, and Risk*. Boulder, CO: Westview Press.

Slovic, P. 1987. Perception of risk. *Science* 236: 280–285.

Slovic, P. 2000. *The Perception of Risk*. London: Earthscan.

Slovic, P. 2010. *The Feeling of Risk: New Perspectives on Risk Perception*. London: Earthscan.

Stern, P.C. and H.V. Fineberg, eds. 1996. *Understanding Risk: Informing Decisions in a Democratic Society*. Report for National Research Council, Committee on Risk Characterization. Washington, DC: National Academy Press.

Svenson, O. 1988. Managing product hazards at Volvo Car Corporation. In *Corporate Management of Health and Safety Hazards: A Comparison of Current Practice*, edited by R.E. Kasperson, J.X. Kasperson, C. Hohenemser, and R.W. Kates, 57–78. Boulder, CO: Westview.

Turner, B.A. 1978. *Man-Made Disasters: The Failure of Foresight*. London: Wykeham.

Tversky, A. and D. Kahneman. 1974. Judgement under uncertainty: Heuristics and biases. *Science* 185: 1124–1131.

Vaughan, D. 1992. Regulating risk: Implications of the Challenger accident. In *Organizations, Uncertainties, and Risk*, edited by J.F. Short, Jr. and L. Clarke, 235–254. Boulder, CO: Westview.

Yong, C. and D.C. Booth. 2011. *The Wenchan Earthquake of 2002*. Beijing: Springer Science Press.

Appendix 8.1

Social amplification of risk: a selective bibliography

1988 Kasperson, R.E., O. Renn, P. Slovic, H. Brown, J. Emel, R. Goble, J.X. Kasperson, and S. Ratick. The social amplification of risk: A conceptual framework. *Risk Analysis* 8 (2): 177–187.

1988 Rayner, S. Modelling through metaphors to maturity: A commentary on Kasperson et al., "The social amplification of risk." *Risk Analysis* 8 (2): 201–204.

1988 Rip, A. Should social amplification of risk be counteracted? *Risk Analysis* 8 (2): 193–197.

1988 Svenson, O. Mental models of risk, communication, and action: Reflections on social amplification of risk. *Risk Analysis* 8 (2): 199–200.

1989 Kasperson, R.E., O. Renn, P. Slovic, J.X. Kasperson, and S. Emani. Social amplification of risk: The media and public response. In *'89, Waste Processing, Transportation, Storage and Disposal, Volume 1. High-level Waste and General Interest,* edited by R.G. Post, 131–135. Tucson, AZ: University of Arizona.

1990 Machlis, G.E. and E.A. Rosa. Desired risk: Broadening the social amplification of risk framework. *Risk Analysis* 10 (1): 161–168.

1991 Kasperson, R.E. and J.X. Kasperson. Hidden hazards. In *Acceptable Evidence: Science and Values in Hazard Management*, edited by D.C. Mayo and R. Hollander, 9–28. Oxford, UK: Oxford University Press.

1991 Renn, O. Risk communication and the social amplification of risk. In *Communicating Risks to the Public: International Perspectives*, edited by R.E. Kasperson and P.J.M. Stallen, 287–324. Dordrecht, the Netherlands: Kluwer.

1992 Kasperson, R.E. The social amplification of risk: Progress in developing an integrative framework of risk. In *Social Theories of Risk,* edited by S. Krimsky and D. Golding, 153–178. New York: Praeger.

1992 Renn, O., W. Burns, J.X. Kasperson, R.E. Kasperson and P. Slovic. The social amplification of risk: Theoretical foundations and empirical applications. *Journal of Social Issues* 48 (4): 137–160.

1993 Burns, W.J., P. Slovic, R.E. Kasperson, J.X. Kasperson, O. Renn and S. Emani. Incorporating structural models into research on the social amplification of risk: Implications for theory construction and decisionmaking. *Risk Analysis* 13 (6): 611–623.

1996 Kasperson, R.E. and J.X. Kasperson. The social amplification and attenuation of risk. *Annals of the American Academy of Political and Social Sciences* 545 (May): 95–105.

1996 Metz, W.C. Historical application of a social amplification of risk model: Economic impacts of risk events at a nuclear weapons facility. *Risk Analysis* 16 (2): 185–193.

1997 Pidgeon, N. *Risk Communication and the Social Amplification of Risk—Phase 1 Scoping Study.* Report to the UK Health and Safety Executive. RSU Ref. 3625/R62076. London: The Executive.

1998 Smith, D. and J. McCloskey. Risk and crisis management in the public sector: Risk communication and the social amplification of public sector risks. *Public Money and Management* 18 (4): 41–50.

1999 Kasperson, R.E. The social attenuation and amplification of risk. In *Risque & société: Actes du Colloque Risque & Société réalisé sous l'égide de l'Académie des Sciences dans le cadre du Centenaire de la Découverte de la Radioactivité, Cité des Sciences et de L'Industrie de Paris-LaVillette, 18–19–20 Novembre 1998,* edited by Maurice Tubiana, Constantin Vrousos, Catherine Carde, and Jean-Pierre Pagès, 111–122. Gif-sur-Yvette, France: Editions Nucléon.

1999 Pidgeon, N. Risk communication and the social amplification of risk: Theory, evidence, and policy implications. *Risk, Decision and Policy* 4 (1): 1–15.

2001 Hill. A. Media risks: The social amplification of risk and the media debate. *Journal of Risk Research* 4: 209–226.

2001 Kasperson, J.X. and R.E. Kasperson. Transboundary risks and social amplification. In *Cross-national Studies of Transboundary Risk Problems,* edited by J. Linnerooth-Bayer and R. Löfstedt, 207–243. London: Earthscan.

2001 Kasperson, R.E., N. Jhaveri and J.X. Kasperson. Stigma and the social amplification of risk. In *Risk, Media, and Stigma: Understanding Public Challenges to Modern Science and Technology*, edited by J. Flynn, P. Slovic, and H. Kunreuther, 9–27. London: Earthscan.

2001 Petts, J., T. Horlick-Jones, and G. Murdock. *Social Amplification of Risk: The Media and the Public.* Health and Safety Executive Contract Research Report 329/2001. Sudbury, UK: HSE Books.

2001 Wåhlberg, A. The theoretical features of some current approaches to risk perception. *Journal of Risk Research* 4 (3): 237–250.

2003 Breakwell, G.M., and J. Barnett. Social amplification and the layering method. In *The Social Amplification of Risk*, edited by R.E. Kasperson and P. Slovic, 80–101. Cambridge, UK: Cambridge University Press.

2003 Freudenburg, W.R. Institutional failure and the organizational amplification of risks: The need for a closer look. In *The Social Amplification of Risk,* edited by N. Pidgeon, R.E. Kasperson and P. Slovic, 102–120. Cambridge, UK: Cambridge University Press.

2003 Frewer, L.J. Trust, transparency, and social context: Implications for social amplification of risk. In *The Social Amplification of Risk,* edited by N. Pidgeon, R.E. Kasperson and P. Slovic, 123–137. Cambridge, UK: Cambridge University Press.

2003 Gowda, M.V. Integrating politics with the social amplification of risk framework: Insights from an exploration in the criminal justice contest. In *The Social Amplification of Risk,* edited by N. Pidgeon, R.E. Kasperson and P. Slovic, 305–325. Cambridge, UK: Cambridge University Press.

2003 Kasperson, J.X., R.E. Kasperson, N. Pidgeon and P. Slovic. The social amplification of risk: Assessing fifteen years of research and theory. In *The Social Amplification of Risk*, edited by N. Pidgeon, R.E. Kasperson and P. Slovic, 13–46. Cambridge, UK: Cambridge University Press.

2003 Leiss. W. Searching for the policy relevance of the risk amplification framework. In *The Social Amplification of Risk,* edited by N. Pidgeon, R.E. Kasperson and P. Slovic, 355–375. Cambridge, UK: Cambridge University Press.

2003 MacGregor, D.G. Public response to Y2K: Social amplification and risk adaptation: Or "how I learned to stop worrying and love Y2K." In *The Social Amplification of Risk,* edited by N. Pidgeon, R.E. Kasperson and P. Slovic, 243–261. Cambridge, UK: Cambridge University Press.

2003 Murdock, G., Pens, J., and T. Horlick-Jones. After amplification: Rethinking the role of the media in risk communication. In *The Social Amplification of Risk,* edited by N. Pidgeon, R.E. Kasperson and P. Slovic, 156–178. Cambridge, UK: Cambridge University Press.

2003 Pidgeon, N., R.E. Kasperson, and P. Slovic, eds. *The Social Amplification of Risk,* Cambridge, UK: Cambridge University Press.

2003 Pidgeon, N., R.E. Kasperson, and P. Slovic. Introduction. In *The Social Amplification of Risk,* edited by N. Pidgeon, R.E. Kasperson and P. Slovic, 1–12. Cambridge, UK: Cambridge University Press.

2003 Poumadere, M. and G. Mays. The dynamics of risk amplification and attenuation in context: A French case study. In *The Social Amplification of Risk,* edited by N. Pidgeon, R.E. Kasperson and P. Slovic, 209–242. Cambridge, UK: Cambridge University Press.

2003 Renn, O. Social amplification of risk in participation: Two case studies. In *The Social Amplification of Risk,* edited by N. Pidgeon, R.E. Kasperson and P. Slovic, 374–401. Cambridge, UK: Cambridge University Press.

2003 Rosa, E.A. The logical structure of the social amplification of risk framework (SARF): Metatheoretical foundations and policy implications. In *The Social Amplification of Risk*, edited by N. Pidgeon, R.E. Kasperson and P. Slovic, 47–79. Cambridge, UK: Cambridge University Press.

2003 Susarla, A. Plague and arsenic: Assignment of blame in the mass media and the social attenuation and amplification of risk. In *The Social Amplification of Risk,* edited by N. Pidgeon, R.E. Kasperson and P. Slovic, 179–208. Cambridge, UK: Cambridge University Press.
2003 Wiedemann, P.M., M. Clauberg, and H. Schiitz. Understanding amplification of complex risk issues: The risk story model applied to the EMF case. In *The Social Amplification of Risk,* edited by N. Pidgeon, R.E. Kasperson and P. Slovic, 286–304. Cambridge, UK: Cambridge University Press.
2012 Burgess, A. Media, risk, and absence of blame for "acts of God": Attenuation of the European volcanic ash cloud of 2010. *Risk Analysis* 32 (10): 1693–1702.
2012 Kasperson, R.E. A perspective on the social amplification of risk. *The Bridge* 42 (3): 23–27.
2012 Kasperson, R.E. The social amplification of risk and low-level radiation. *Bulletin of Atomic Scientists* 68 (3): 59–65.
2013 Kasperson, R.E. Risk governance and the social amplification of risk: A commentary. In *The Articulation of Hazard, Politics, and Ecology*, edited by U. Fra, 485–487. Berlin: Springer.

Part 4

Case studies of extreme risks

9 Safety and severe accidents in nuclear reactors

Michael Corradini and Vicki Bier

Philosophy of nuclear power safety

Safety overview

The main safety concern in the operation of nuclear power plants and their associated fuel fabrication facilities is the emission of uncontrolled radiation releases into the environment. By their very nature, nuclear power plants use nuclear fuel (e.g., uranium in an oxide form) in the fission process to produce heat and thereby electrical power. Thus, the nuclear plant contains the largest inventory of radioactive materials that must be safeguarded. Accidents need to be minimized and nuclear fuel cooled, so that it does not become damaged, overheat, and possibly melt, releasing radioactive materials.

Maintaining the public's health and safety while allowing for the continued development and utilization of nuclear energy has been a major focus in both Europe and the U.S., as well as in many other countries in which nuclear energy is used for power production. One of the difficulties with public acceptance of nuclear power is that safety is a difficult concept to define and, once defined, is not easily demonstrated to the general public. In the aftermath of the recent accident at Fukushima, people have legitimate questions about how the nuclear power industry and regulators can ensure nuclear safety.

To complicate matters, the public may have concerns that transcend considerations of nuclear power safety; e.g., effects of radiation on public health in medical applications. The public is still relatively unfamiliar with nuclear power technology, and thus is prone to judge it critically and be cautious in accepting it as a useful and valuable energy source. Often, the only familiarity the public has with nuclear power is through media coverage of severe accidents (Three Mile Island, Chernobyl, and Fukushima; see Table 9.1) or near misses (e.g., Browns Ferry, Davis-Besse[1]). Thus, public fears of nuclear power are maintained by the belief that catastrophic consequences loom as a hazard for the general public if an accident occurs. The Chernobyl accident in particular reinforced these perceptions, even though the reactor design was atypical of any plant outside the USSR. The historic earthquake and tsunami that initiated the Fukushima Daiichi reactor accident again reminded safety engineers of the need to consider the potential causes of severe accidents, regardless of their likelihood or origin. Finally, because these possible accident hazards affect the general public, there is concern due to the involuntary nature of the risk to which people are exposed.

In this chapter, we do not focus primarily on public perceptions of the risks of nuclear power, but instead emphasize how the nuclear industry and regulatory agencies model, analyze, and manage the risks of nuclear power. Fundamentally, nuclear power safety is achieved by the reliable operation of the nuclear power plant, with a minimum number of abnormal events. This is a good starting point for defining safety because it points out that safety is rooted in proper design and operation of the technology, which minimizes the risk of accidents. However, one cannot focus entirely on reliable operation and accident prevention. The natural complement to reliability and prevention is protection of the public by mitigating the consequences of an accident, if one occurs.

This combination of prevention through reliable operation and mitigation may seem straightforward, but nuclear power has two attributes that make this a difficult challenge. First, all technologies rely to some degree on the ability to empirically test that a design or engineering system is "safe." Such accident testing is usually done under full-scale conditions where a prototype might be destructively tested; e.g., automobile crash testing for passenger safety. However, full-scale experimentation of nuclear power accidents is technically and economically impractical. (Even in the commercial airline industry, where the cost of testing is much less, there has only been one destructive crash test of a full-scale airliner in decades.) In addition, accidents that result in release of radioactivity into the environment are fortunately infrequent events. Three Mile Island (TMI), Chernobyl, and Fukushima are the only severe commercial reactor accidents that have occurred over the last five decades, despite the industry having accumulated more than sixteen thousand reactor years of commercial experience worldwide. This makes it difficult for the industry to learn about safety from empirical accident experience. As a result, the industry relies on the concept of "defense in depth" to ensure adequate safety.

Defense in depth

The defense-in-depth concept (which is used in nuclear programs around the world) relies on three diverse and independent approaches to achieve safety. The first line of defense emphasizes the prevention of any abnormal events that could lead to damage to the nuclear fuel, radioactivity release, and thereby public risk. The second line of defense consists of specific plant engineering features that are intended to prevent damage to the nuclear fuel in case of certain categories of abnormal events or accidents (referred to as the "design basis" of the plant). Such events might include, for example, a pipe break in the plant, or an earthquake of moderate intensity. Despite these events being quite unlikely and obviously not part of normal operation, nuclear plants are designed to be able to withstand these "design basis accidents" (DBA), without damage to the nuclear fuel leading to core melt. Finally, the third aspect of defense in depth involves plant features that have no role in normal operation, but whose sole purpose is to prevent the release of radioactivity into the environment in the event of a severe accident such as TMI. The simplest example of such features is the containment building itself; if one could be assured that a severe accident causing damage to the nuclear fuel would never happen, there would be no need for a containment building.

Commercial light water reactor severe accidents

Three Mile Island Unit 2 (TMI) nuclear power plant (Pennsylvania, U.S.)

TMI Unit 2 was a pressurized water reactor (PWR). The plant began operation in December 1978 and suffered a core-melt accident in March 1979. The damage from the accident was severe enough that the reactor was decommissioned. The nuclear fuel damage resulted in a small amount of radioactivity being released to the environment.

Chernobyl Unit 4 nuclear power plant (Ukraine, former Soviet Union)

Chernobyl Unit 4 was a graphite-moderated boiling water reactor (BWR). The Chernobyl nuclear power plants were originally commissioned in 1972. Unit 4 was destroyed in a core-melt accident in April 1986. The damage released over 50 million curies of radioactive materials and caused 50 fatalities due to fire and acute radiation exposure.

Fukushima Daiichi site (Fukushima Prefecture, Japan)

Unit 1 BWR
Unit 1 was connected to the electricity grid in November 1970 and began commercial operation in March 1971.

Unit 2 BWR
Unit 2 was connected to the electricity grid in December 1973 and began commercial operation in July 1974.

Unit 3 BWR
Unit 3 was connected to the electricity grid in February 1978 and began commercial operation in October 1978.

Units 1, 2, and 3 all suffered core melts following an earthquake in March 2011. The extent of damage is still unknown, as decommissioning efforts are still underway. Based on measurements and analysis, however, the amount of radioactivity released was about an order of magnitude less than what was released at Chernobyl. There were no early fatalities from the accident, but two nuclear reactor operators drowned from the tsunami.

Another way to consider defense in depth is to identify the physical barriers to release of radioactivity, which correspond roughly to the three lines of defense discussed above:

1 Fuel rod cladding keeps all radioactivity enclosed under normal conditions.
2 Engineered safety systems (ESS), such as pumps and valves to deliver cooling water to the core, prevent fuel overheating and cladding failure in the event of a DBA.
3 Containment (and various safety systems to protect the containment building) prevents radioactivity release in the case of beyond-design-basis accidents in which fuel rods overheat, cladding fails, fuel can melt, and radioactivity is no longer contained within the fuel rods.

As stated in the original Reactor Safety Study conducted by the U.S. Nuclear Regulatory Commission (NRC 1975, WASH-1400), the performance of these multiple lines of defense should be minimally dependent upon each other and should be continually assessed using reliability analysis as well as deterministic engineering calculations.

Brief history of reactor safety and risk

As with any technological development, the historical evolution of nuclear power safety did not follow a simple path. In fact, ever since the 1950s, it was recognized that the only significant risk to the public from a nuclear power plant was the chance of an accident in which nuclear fuel damage and possibly core melt occurred, and the radioactive fission products were released to the environment. At that time, the primary way of ensuring safety in case of an accident was to site nuclear plants in remote regions of the countryside, so that the large distance from the plant to population centers would minimize the risk to the public following atmospheric dispersion of any released fission products (Kerr 1989). This approach was used in most countries, not only for large commercial nuclear plants but also for smaller research reactors. Moreover, safety analysis generally assumed (conservatively) that, in the event of a severe accident, the complete radioactive inventory of the reactor core would be released. This type of bounding analysis minimized the need for detailed research on possible accident sequences, so research initially focused on development of models for human health effects in the event of a radiation release, leading to highly conservative hazard analysis by the U.S. Atomic Energy Commission (AEC 1957), and ultimately the establishment by the AEC of formal criteria requiring remote siting of nuclear plants in the Code of Federal Regulations (10CFR100) and technical information document (TID) 14844 (DiNunno et al. 1962). Similar siting criteria were adopted in some fashion by many other regulatory authorities. These siting criteria assumed that the "maximum credible accident" would be a core melt, and specified an arbitrary assumption regarding how much radiation would be released (the "source term"), rather than relying on detailed calculations of source terms.

In the 1960s, nuclear power plant design and construction was the primary activity of the industry, so safety analysis focused on the ability of the engineering systems to prevent abnormal occurrences and to cope with a DBA. As the power levels of light water reactors were increased (to take advantage of economies of scale), in order to maintain adequate core cooling under accident conditions, development focused on enhancing the design of the engineered safety systems (ESS) and the containment system, to accommodate specific DBA scenarios (Okrent 1981). Such DBAs might involve a loss of cooling in the reactor core (caused by a pump failure or a pipe break), or a temporary overpower condition in the reactor core. These issues were originally intended to be resolved by detailed calculations and comparison to large-scale tests at the Idaho National Engineering Laboratory. However, due to experimental delays and the need for immediate regulatory guidelines, an interim criterion was instituted in 10CFR50 (Appendix K)[2] after regulatory hearings in the early 1970s; this document specified both acceptable design features and the accepted methods of analysis. After adoption of these regulatory guidelines, research was initiated in order to improve understanding of the thermal hydraulics governing reactor behavior in these DBAs, and to confirm that the approved ESS designs were sufficiently "conservative," or safe. Thus, the major emphasis of safety work in the 1960s and mid-1970s focused on handling DBA scenarios, although there were also continued efforts to improve the containment system and the concept of defense in depth.

In the 1960s, it had been recognized that the "maximum credible accident" might be neither "maximum" nor "credible." In particular, one could not evaluate whether a postulated severe accident was "credible" without some consideration of its likelihood; moreover, as one moved to less and less likely scenarios, the consequences might become more severe than was assumed in the "maximum credible accident." In response, Farmer (1967) suggested a probability-based technique for hazard analysis of nuclear analysis, adapting an approach that was then beginning to be used in the British aviation industry. This "probabilistic reliability analysis" technique, or "probabilistic risk assessment" (PRA, discussed in more detail in Chapter 1 in this volume), was fundamentally different than the deterministic methods of accident analysis that had been applied to DBAs, because it focused not only on accident consequences, but also on estimating the probabilities of various accident scenarios. In other words, DBA analyses focused on postulating a particular accident sequence and deterministically analyzing its consequences. By contrast, PRA considers a myriad of possible accident sequences, each with an associated probability of occurrence; the overall risk then reflects the probabilities and consequences of all of these possible accident sequences.

In the early 1970s, the NRC (the successor agency to the AEC) applied this new technique to two operating reactors—Surry (a PWR) and Peach Bottom (a BWR)—to estimate the safety risk to the public, and also the sources of the risk. This analysis was documented in the so-called Reactor Safety Study (NRC 1975, WASH-1400). This technique was quite novel in the reactor safety community. While the Reactor Safety Study admittedly had some weaknesses in terms of completeness and the treatment of uncertainties, it did provide a benchmark for reactor safety and established a standard methodology that is still used today. This study also developed some important insights for beyond-design-basis accidents, as part of an integrated analysis to determine the risk from nuclear power plants:

- Accidents that resulted in fuel rod failure and core melt dominated the risk.
- Core-melt accidents were estimated to be of low probability but high consequence.
- Probability of core melt was in the order of one in 10,000 per reactor per year.
- Consequences might be quite severe for certain accident sequences.
- The risk to the public was small in comparison to everyday man-made risks.

While the WASH-1400 assessment was not directly used in NRC licensing decisions, some limited research was initiated in Britain, Germany, and the U.S. on the key physical processes involved in core melt (Butcher 1975; Dahlgren 1976). This reflected an advance over the prior small level of effort invested in testing and analysis for severe accidents.

Despite the development of PRA, the accepted method of safety analysis in the 1970s was still based on defense in depth, with deterministic analyses used to analyze postulated abnormal occurrences and DBA scenarios and assess the performance of the containment system. This approach was thought to be sufficiently conservative that the actual hazard to the public would be insignificant. However, the TMI accident in 1979 altered this perception. The TMI accident involved a series of equipment malfunctions and operator actions that caused an abnormal occurrence to evolve into a severe accident over a four-hour period; a significant fraction of fuel melted, so radioactivity was released from the reactor core. Although radioactive material was released from the melted fuel, the defense-in-depth strategy of containment successfully prevented any major release of radioactivity into the environment. However, the TMI accident did raise doubts about the soundness of using a completely deterministic approach to safety analysis, and also about the lack of detailed understanding of the physical processes that could threaten containment integrity. These issues became a major focus of safety investigations in the 1980s in the international nuclear safety community, continuing to this day. The accident at Chernobyl in 1986 reinforced these efforts, even though it involved a unique reactor design that has not been used outside the USSR. Specific insights from the TMI and Chernobyl accidents for reactor safety risk are noted below.

Severe accident findings for TMI-2 and Chernobyl

TMI-2 accident findings

- The accident at TMI fell in between a DBA and a complete core melt in severity.
- It released a small amount of radioactivity (less than 10 curies of iodine).
- Offsite consequences were small (well within the bounds specified in 10CFR100).
- The accident evolved over hours, with operator actions mitigating the effects.
- The accident was a significant financial burden to the utility ($2 billion in costs).
- The NRC began to rethink how it should consider the risk from nuclear plants:
    - Physical processes during a core melt may reduce the size of a release.
    - Strong containments and other passive safety features were more effective than previously believed.

Chernobyl accident findings

- The accident at Chernobyl was significantly more severe than TMI.
- It was financially and environmentally disastrous for the USSR.
- It released a large amount of radioactivity into the environment.
- Offsite consequences were large (although lower than anticipated).
- The accident evolved too quickly for human actions to mitigate the consequences.
- Chernobyl caused the world community to reconsider the Chernobyl-type design for use in electrical power production.

Nuclear safety goals

Because of the TMI accident, the reactor safety community in the U.S. (as well as in other countries) reexamined its reactor safety regulations and the difficult question of how safe was safe enough. This consideration was most pronounced in the U.S., since TMI was a U.S. plant; see, for example, President's Commission on the Accident at TMI (1979), the Electric Power Research Institute (EPRI 1979), and NRC (1980).

As a result of these investigations, two activities were initiated that still continue today. First, the private nuclear industry formed a consortium, Industry Degraded Core Rulemaking, to pool its resources to research the processes involved in severe reactor accidents and develop analytical tools to predict accident progression and effects. This function for industry was eventually incorporated into the EPRI nuclear safety program, and now has international partners.

Second, the NRC proposed a set of quantitative safety goals for nuclear power plants in the 1980s, to ensure that the risks from nuclear power remained small relative to other risks to the public. These draft safety goals, published in 1986 and entitled "Safety Goals for the Operation of Nuclear Power Plants; Policy Statement," were as follows:

- The risk to an average individual in the vicinity of a nuclear power plant of prompt fatalities that might result from reactor accidents should not exceed 0.1 percent of the sum of prompt fatality risks resulting from other accidents to which members of the U.S. population are generally exposed.
- The risk to an average individual in the vicinity of a nuclear power plant of cancer fatalities that might result from reactor accidents should not exceed 0.1 percent of the sum of cancer fatality risks resulting from all other causes.

Subsequent to this draft safety-goal statement, the NRC over the next fifteen years amended the safety goals and refined their application in regulatory usage.[3]

These safety goals have been supplemented by surrogate numerical risk goals that are easier to evaluate. These surrogate numerical risk goals are: (1) that the likelihood of core melt should be less than one in 10,000 per reactor year; and (2) that a large release of radioactive materials should be at least a factor of ten less likely than a core melt.

Severe accident initiation

A severe accident can occur in a nuclear power plant if the reactor system experiences some initiating event that disturbs normal operation, followed by multiple failures of safety systems. Such an initiating event could involve a loss of reactor core cooling caused by a pump failure, pipe break, or loss of electric power, or an overpower incident. An initiating event by itself would not be sufficient to cause a severe accident, because of ESS features that are designed to activate automatically to provide backup cooling, power control, or shutdown capabilities. In addition, each safety system is designed with redundant components (e.g., more pumps than are strictly needed to provide cooling), so that any safety system can still perform its function successfully in the event of a single component failure.

Safety system failures can occur due to a malfunction of one or more components of the system (e.g., pumps, valves, electrical relays), due to improper operator action (e.g., an operator incorrectly shutting off a pump or closing a valve), or due to an "external event" (e.g., a flood or earthquake). External events such as these can be especially problematic, since they can cause failures of multiple components and systems.

To summarize, a severe accident would need to involve a combination of an initiating event and one or more system failures (not just component failures); e.g.:

1 a small pipe break combined with failure of high pressure water injection; or
2 a reactor shutdown combined with an extended loss of alternating current (AC) electrical power.

PRA generally analyzes such postulated severe accident scenarios in three stages. The first stage involves identifying all possible accident sequences that could lead to core melt and quantifying how likely they are. Challenges in this stage of PRA involve developing rigorous models of human reliability, quantifying the likelihood of external events such as earthquakes, and conducting uncertainty analysis of the results. For examples of risk analyses that include these types of models, consider NUREG-1150 (NRC 1990), the

German Risk Study (Global Research for Safety 1990), and the Sizewell B PRA in the United Kingdom (Westinghouse Electric Corporation1992). This aspect of PRA has become a mainstay of nuclear-plant design and operation, and is frequently used to identify design and operational improvements.

Assuming that a postulated severe accident continues all the way to core melt, then the second stage of the PRA would be a containment failure analysis. The purpose of this stage is to estimate the probability and severity of containment failure, either due to physical processes that exceed the strength of the containment system or due to operator action that results in the containment being open to the environment. This aspect of PRA is fundamentally different than the previous one because the uncertainty in quantifying containment reliability is dominated by a lack of complete knowledge of the physical processes involved in beyond-design-basis accidents.

The final stage of a PRA is the analysis of offsite consequences if the containment fails. This stage involves estimating the health effects to the public as a result of the leakage of radioactive gases and other materials from the containment. This stage is not discussed further in this paper but is still an important aspect of PRA.

Given a core melt, the pressure and temperature in the containment increase due to physical processes such as hydrogen combustion (as occurred at Fukushima), direct heating, and interactions between the melted fuel and the coolant or concrete in the reactor. The resulting temperature and pressure increases can challenge the integrity of the containment. A number of studies have attempted to systematically identify and quantify these containment pressure and temperature loads, and the likelihood of containment failure as a function of the amount of time elapsed since core melt. For example, the NRC Zion-Indian Point Study (NRC 1981, NUREG-0850) and the Reactor Risk Reference Document (NRC 1990, NUREG-1150) contain a thorough study of containment temperature and pressure loads, and failure modes. NUREG-1150 considered both types of reactors in commercial use in the U.S. (PWR and BWR), as well as multiple different types of containment designs, as illustrated in Figures 9.1 and 9.2.

Despite the existence of different containment designs, all designs rely on the overall safety design philosophy of defense in depth—relying first on fuel cladding, then on the reactor coolant system and additional ESS features, and ultimately on the containment to prevent releases of radioactivity into the environment. Each of the containment designs shown in Figures 9.1 and 9.2 has a method for reducing pressure loadings in the event of an accident, to protect containment integrity. The PWR large dry containment reduces the risk of high containment pressures just by the large volume of the containment, whereas the others use water or ice to remove heat from the gases in the containment, and thereby prevent excessive containment pressures; the BWR designs use water in a pressure-suppression pool (shown at the bottom of the illustrations in Figure 9.2), while the PWR ice condenser (a relatively rare design) uses ice on the outer walls of the containment. In addition, after TMI, the BWR Mark I designs were modified to allow deliberate venting of containment.

In any reactor design, melted fuel from the reactor core in the event of an accident will accumulate in the core and will eventually exit the bottom of the reactor pressure vessel into a relatively small containment compartment labeled the "reactor cavity" in Figure 9.1a and the "pedestal" in Figure 9.2. The BWR Mark I and II designs also use an inert atmosphere to preclude the burning of any combustible gas generated during an accident. Depending on the specific nature of the accident scenario being considered, the melted fuel can generate steam and other gases that increase the pressure in the containment.

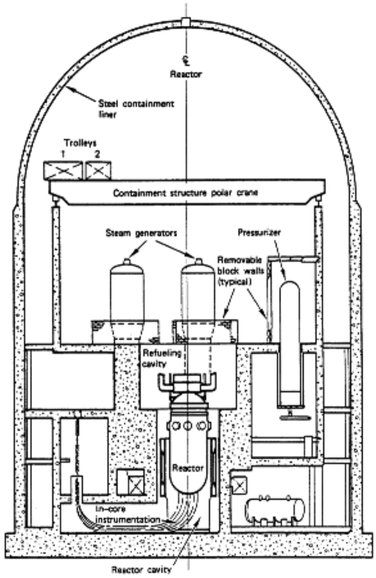


Figure 9.1a PWR large dry containment (as produced in Fukushima Daiichi: ANS Committee Report, 2012).

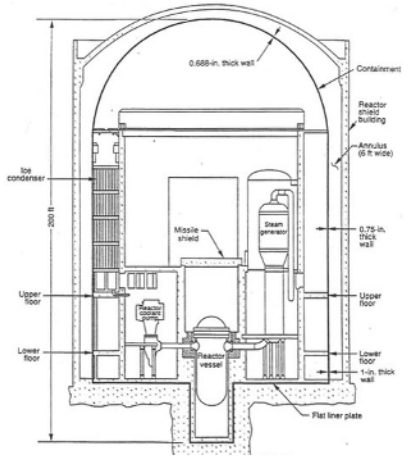


Figure 9.1b PWR ice-condenser containment (as produced in Fukushima Daiichi: ANS Committee Report, 2012).

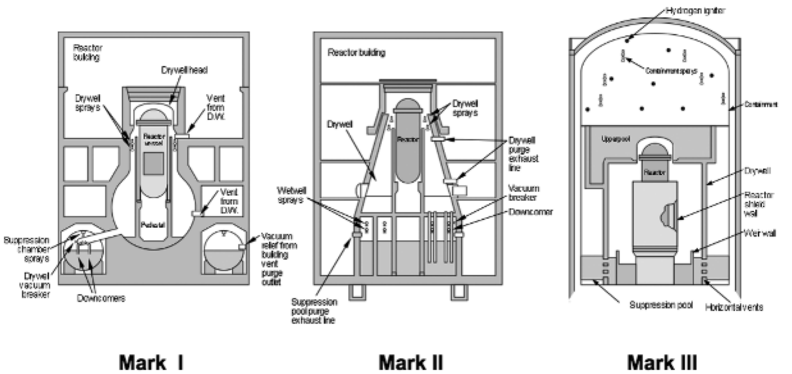


Figure 9.2 BWR containments (as produced in Fukushima Daiichi: ANS Committee Report, 2012).

There is quite a bit of variability between reactor designs, both in the volume of the containment (which determines how much gas it can contain before failing) and how the resulting pressure and temperature loads are managed. For example, the BWR Mark I has a relatively small containment volume, requiring active pressure suppression (as well as operator action to protect the containment under certain circumstances). By contrast, in the PWR large dry containment, pressure suppression is passive, since any gases generated from the melted fuel can just expand into the large volume of the containment.

In the following sections, we use the TMI accident as an example to illustrate some of the most important severe accident phenomena and how they affect containment pressure and temperature. Many of these processes also occurred in the Fukushima accident, but since the investigation of that accident is still underway and the accident is not yet thoroughly understood, we use TMI for illustration purposes. However, we will note that the phenomenon of hydrogen generation and combustion that occurred at Fukushima was also observed at TMI, and has been extensively studied since then; see Sherman et al. (1980) and Breitung et al. (2005). Thus, the hydrogen explosions (not nuclear explosions) that occurred at Fukushima do not represent a previously unanticipated risk.

In the next section, we illustrate some of the phenomena that can happen after the core has become degraded, based on experiences with the TMI-2 accident. (Of course, these results are only illustrative, since we do not go into great technical detail here.) In a later section, we then discuss the recent developments following the Fukushima accident (American Nuclear Society 2012), as a result of which new requirements are being promulgated to address beyond-design-basis accidents.

Severe accident physics—TMI

In the prolonged absence of normal or emergency cooling, the reactor fuel and cladding would eventually become severely distorted due to overheating and melting. This is referred to as a "degraded core." (Note that even if the fuel is not completely melted, a degraded core still poses a problem, since distortions of the fuel rods may interfere with coolant flow.) To summarize the physical insights into this process, we use what was determined to have occurred during the TMI severe reactor accident as an example.

In the absence of sufficient cooling, the fuel will eventually overheat, with the central part of the reactor core melting first (see Figure 9.3 for an illustration). Melted reactor fuel and (nonradioactive) control rods will eventually collect at the bottom of the reactor vessel (see Figure 9.4) and form a pool of material. If the molten mass encounters water at the bottom of the reactor vessel (Figure 9.5), the melted fuel could either experience gradual cooling or else be quenched by mixing with and boiling away the residual water.

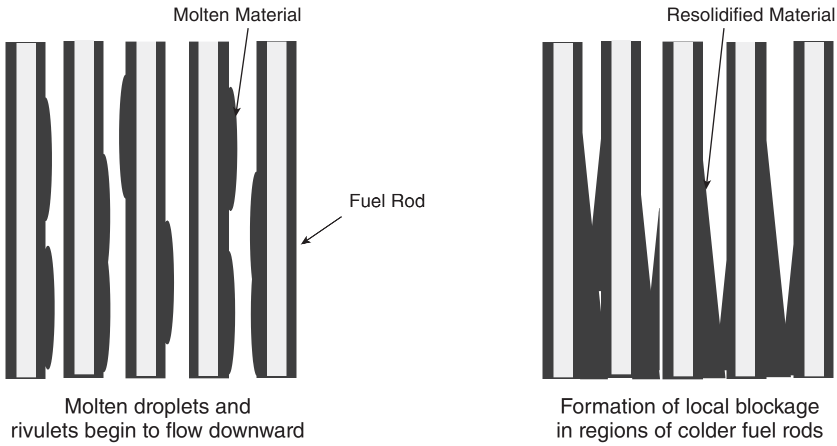


Figure 9.3 Conceptual picture of fuel rod meltdown (as produced in Fukushima Daiichi: ANS Committee Report, 2012).

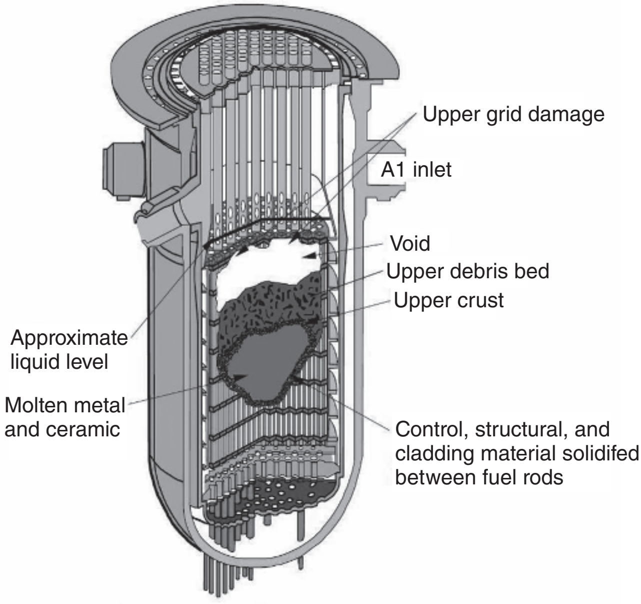


Figure 9.4 Conceptual picture of molten pool formation (as produced in Fukushima Daiichi: ANS Committee Report, 2012).

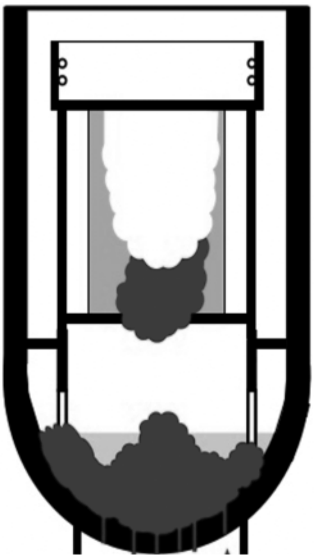


Figure 9.5a Molten core slump/pour (as produced in Fukushima Daiichi: ANS Committee Report, 2012).

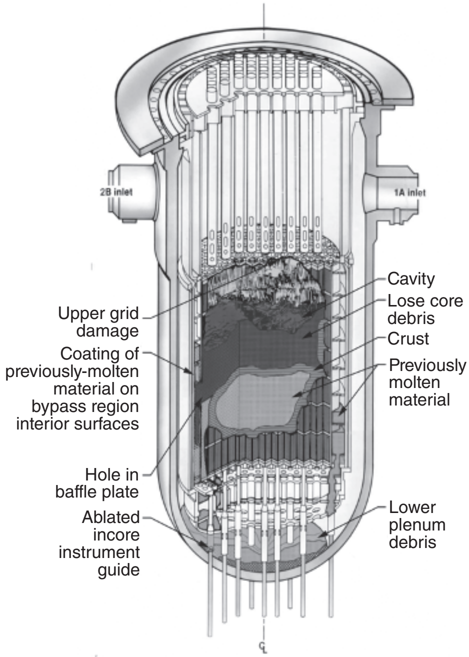


Figure 9.5b TMI-2 end-state configuration (as produced in Fukushima Daiichi: ANS Committee Report, 2012).

Based on TMI investigations, when the core-melt material encounters water at the bottom of the reactor vessel, it will eventually cool, solidify, and form a bed of "debris." At TMI, it is estimated that about 20 metric tons of core materials relocated to the lower part of the reactor vessel, cooling and eventually forming a hard layer. The debris was cooled by the remaining water in the TMI core, and the severe accident was arrested.

Information made available by post-accident investigations (Olsen et al. 1989) shows that a large part of the core was disrupted, resulting in the formation of a "cavern" or vacant area approximately 1.5 m deep in the reactor core, a debris bed below this cavity, and a solid mass of previously melted material below that (see Figure 9.5b). However, we know that this material was successfully kept cool, since the bottom of the reactor vessel did not fail. Information from this event (e.g., about the temperatures that were achieved and the length of time that those temperatures were maintained) is extremely useful in gaining insight into the physics of severe accidents.

To summarize, in the TMI-2 accident, we know that: (1) the integrity of the reactor vessel was maintained, even though approximately a quarter of the core material slumped down to the lower part of the reactor vessel; and (2) a relatively small amount of coolant was enough to cool the melted material. As a result, little radioactivity left the containment; most radioactive material was retained in the reactor vessel. This is overall encouraging, since it suggests that even a severely degraded core can be cooled before it breaches the reactor vessel. Thus, lessons learned from TMI involve the need to:

1. flood the core with water if available, to halt further core melting; and
2. quench the melted core materials at the bottom of the reactor vessel.

Effective methods of achieving these goals have been the topic of ongoing research work.

Mitigation of beyond-design-basis accidents—Fukushima

The March 2011 Tohoku earthquake and resulting tsunami exceeded the design basis for the Fukushima Daiichi site and the level of protection required for these nuclear plants. The Jogan tsunami in 869 AD and related tsunamis suggest a return period of approximately one thousand years for such large tsunamis. Additional protections would have been necessary to protect against these extreme external events; e.g., a higher seawall to protect against tsunami surges, waterproof compartments for emergency power sources, or positioning emergency equipment at higher elevations. This discrepancy between the design basis and the historical record emphasizes the need for a coherent risk-informed regulatory approach, utilizing the most advanced evaluation methodologies and accounting for all relevant available data. This was not done for the Fukushima Daiichi site.

It has long been recognized in PRAs that external events, particularly seismic and external flooding events, could be substantial contributors to risk because of the potential that such extreme events could go beyond the design-basis events in a plant design. This has led to regulatory actions that seek to mitigate beyond-design-basis accidents.

Regulatory actions for beyond-design-basis events

Only a few weeks after the Fukushima Daiichi reactors and spent fuel pools were stabilized by efforts of the Japanese utilities and government, with assistance from other groups, many nations and international bodies began to review their nuclear power plant performance and assess current safety measures. In the United States, two major efforts were begun. The U.S. NRC appointed a six-member Near-Term Task Force (NTTF). In addition, the U.S. nuclear industry, led by the Institute of Nuclear Power Operations and the Nuclear Energy Institute, worked with nuclear-plant owner–operators to review their safety posture and develop a set of lessons learned that would help prevent such an event in the U.S.

The NTTF (NRC 2011) found that even in light of the events at Fukushima Daiichi, current nuclear-plant operation posed no safety hazards, and current safety measures provided adequate protection to public health and safety. However, the NTTF also found that while the defense-in-depth philosophy (enhanced by PRA) continues to serve as the primary organizing principle of regulation, the defense-in-depth safety approach could be strengthened by including explicit requirements for beyond-design-basis events.

In particular, the NTTF recommended a series of twelve broad actions to improve safety in the event of extreme natural disasters and other associated low-probability events, such as loss of electric power. The NRC used these recommendations as the basis to develop a series of regulatory actions for the U.S. nuclear industry. By the first anniversary of the Fukushima accident, these regulatory actions had taken the form of immediate requests for information from nuclear-plant owner–operators, orders for immediate actions by nuclear-plant owners, and regulatory rulemaking.[4] These actions included reevaluation of the seismic and flooding hazards at U.S. nuclear plants, as well as reviews of current protections against extreme natural disasters. In addition, the NRC, in collaboration with the industry, developed a series of mitigation strategies for beyond-design-basis events, the so-called FLEX program.

The FLEX mitigation approach

The NRC has approved, and the industry has adopted, a set of emergency equipment to cope with a range of beyond-design-basis events that may be initiated by either equipment failures or natural disasters. The objective is to establish diverse, redundant, and independent resources (equipment, personnel, and associated procedures) to provide flexible coping capabilities to prevent damage to the fuel in the reactor and spent fuel pools, and maintain containment integrity, by using installed equipment, on-site portable equipment, and pre-staged offsite equipment. These resources include:

- portable equipment to assure that multiple means of obtaining power and cooling water are available to support key safety functions for all reactors at a site; equipment includes portable pumps, generators, batteries, battery chargers, compressors, hoses, couplings, tools, debris clearing equipment;
- protection of portable equipment to guard it from the severe natural phenomena predicted for that site by locating the equipment in safe locations;
- procedures and guidance for emergency response personnel for the use of FLEX equipment and capabilities;
- program controls to ensure regular maintenance and testing of FLEX equipment;
- training on FLEX capabilities and responses.

The rationale for FLEX is that if this set of equipment is kept ready in a safe location, there would be working equipment available in the event of an emergency, no matter the disaster that befalls the facility. Therefore, the risk to the public is significantly reduced.

Conclusions

To conclude, nuclear power safety is best achieved by the reliable operation of the nuclear power plant with a minimum number of abnormal events. Nevertheless, the concept of defense in depth requires that one does not focus exclusively on reliable operation and accident prevention, but also designs appropriate systems (i.e., emergency core-cooling systems and the containment system) to maintain core cooling and prevent radioactivity release in the event of an accident. Even then, one must consider accidents that exceed the design basis (severe accidents) and use PRA to identify these scenarios, ascertain their likelihood, and develop means to mitigate their consequences; i.e., FLEX. No engineering system is perfectly safe, and nuclear power is no different. However, the combination of risk assessment, mitigation of severe accidents, and learning from historical experience can ensure that these events are rare and unlikely to be catastrophic.

Notes

1 On Browns Ferry see https://www.nrc.gov/reactors/operating/ops-experience/fire-protection.html; on Davis-Besse see https://www.nrc.gov/reactors/operating/ops-experience/vessel-head-degradation.html
2 https://www.nrc.gov/reading-rm/doc-collections/cfr/part050/part050-appk.html
3 These discussions, and links to the original NRC documents on the subject, can be found at https://www.nrc.gov/reading-rm/doc-collections/commission/secys/2000/secy2000-0077/2000-0077scy.pdf
4 All of these actions can be found at http://www.nrc.gov/reactors/operating/ops-experience/japan-dashboard.html

References

Atomic Energy Commission (1957) *Theoretical possibilities and consequences of major accidents in large nuclear power plants*, WASH-740. Atomic Energy Commission, Washington, DC.

American Nuclear Society Special Committee on Fukushima (2012) *Fukushima Daiichi: ANS Committee Report* http://fukushima.ans.org/report/Fukushima_report.pdf

Breitung, W., S. Dorofeev, A. Kotchourko, R. Redlinger, W. Scholtyssek, A. Bentaib, J.-P. L'Heriteau, P. Pailhories, J. Eyink, M. Movahed, K.-G. Petzold, M. Heitsch, V. Alekseev, A. Denkevits, M. Kuznetsov, A. Efimenko, M.V. Okun, T. Huld, and D. Baraldi (2005) "Integral large scale experiments on hydrogen combustion for severe accident code validation-HYCOM," *Nuclear Engineering and Design* 235, 253–270.

Butcher, B. (1975) *Light Water Reactor Safety Research Program Quarterly Report, July–September 1975*, SAND75-0632, Sandia National Laboratories, Albuquerque, NM.

Dahlgren, D. (1976) *Light Water Reactor Safety Research Program Quarterly Report, October–December* 1975, SAND76-0163, Sandia National Laboratories, Albuquerque, NM.

Electric Power Research Institute (1979) *Analysis of Three Mile Island—Unit 2, Nuclear Safety Analysis Center Report*, Electric Power Research Institute, Palo Alto, CA.

DiNunno, J.J., F.D. Anderson, R.E. Baker, and R.L. Waterfield (1962) *Calculation of Distance Factors for Power Reactor Sites*, Technical Information Document TID-14844, Atomic Energy Commission, Washington, DC.

Farmer, F.R. (1967) "Siting Criteria—A New Approach," *Containment and Siting of Nuclear Power Plants: Proceedings of a Symposium on the Containment and Siting of Nuclear Power Plants*, International Atomic Energy Agency, Vienna, Austria, pp. 303–324.

Global Research for Safety (1990) *German Risk Study Nuclear Power Plants, Phase B.* Gesellschaft für Reaktorsicherheit, BMFT, Bonn, Germany.

Kerr, W. (1989) "History of Containment Design Criteria," Topical Meeting on Nuclear Power Plant Thermal Hydraulics and Operations, NUPTHO-4, Seoul, Korea.

Nuclear Regulatory Commission (1975) *Reactor Safety Study: An Assessment of Accident Risks in US Commercial Nuclear Power Plants*, WASH-1400, NUREG-75/014, Nuclear Regulatory Commission, Washington, DC.

Nuclear Regulatory Commission (1980) *Three Mile Island: A Report to the Commissioners and to the Public*, Nuclear Regulatory Commission, Washington, DC.

Nuclear Regulatory Commission (1981) *Preliminary Assessment of Core Melt Accidents at the Zion and Indian Point Nuclear Power Plants and Strategies for Mitigating Their Effects*, NUREG-0850, V1, Nuclear Regulatory Commission, Washington, DC.

Nuclear Regulatory Commission (1990) *Reactor Risk Reference Document*, NUREG-1150. Nuclear Regulatory Commission, Washington, DC.

Nuclear Regulatory Commission (2011) *The Near-Term Task Force Review of Insights from the Fukushima Dai-ichi Accident*, Nuclear Regulatory Commission, Washington, DC. http://pbadupws.nrc.gov/docs/ML1118/ML111861807.pdf

Okrent, D. (1981) *Nuclear Reactor Safety*, University of Wisconsin Press, Madison, WI.

Olsen, C.S., R.R. Hobbins, and B.A. Cook, (1989) "Application of Severe Fuel Damage Experiments to Evaluating Three Mile Island Unit 2 Core Materials Behavior," *Nuclear Technology*, 87, 884.

President's Commission on the Accident at TMI (1979) *The need for change: The legacy of TMI: report of the President's Commission on the Accident at Three Mile Island, John G. Kemeny, chairman*, Washington, DC.

Sherman, M.P., M. Berman, and J.C. Cummings (1980) *The behavior of hydrogen during accidents in LWRs*, SAND80-1495, NUREG/CR-1561, Sandia National Laboratories, Albuquerque, NM.

Westinghouse Electric Corporation (1992) *Sizewell B Probabilistic Safety Study*, WCAP 9991, Westinghouse Commercial Atomic Power, British Energy, London, UK.

10 Mitigating extreme infectious disease disaster risk

Terrence M. O'Sullivan

> Even though we can't compute the odds for threats like bioterrorism or a pandemic, it's important to have the right people worrying about them and taking steps to minimize their likelihood and potential impact. … But bioterrorism and pandemics are the only threats I can foresee that could kill over a billion people.
>
> Microsoft Corporation Chairman, Bill Gates, 2010[1]

Infectious disease outbreaks come in all sizes, shapes, and forms. Each is different: They can be brief or long-lasting, small and deadly, acute or slow-building, large and merely inconvenient, or, in the worst cases, large-scale lethal pandemics that spread easily across continents, infecting hundreds of millions. They can also be naturally occurring or a product of man-made biological warfare or terrorism. Most Western nations are relatively unaccustomed to genuine infectious disease *disasters*, given the effectiveness of modern vaccines, drugs, and basic public health. But events like the 2001 anthrax attacks, 2003 SARS epidemic, 2014 Ebola outbreak, and even the emerging Zika virus have demonstrated that the potential for damaging and/or disruptive outbreaks is rising. This increased risk is tied also to increasingly bleak projections of the impacts of climate change; expanding global trade, travel, and migration; new pathogenic[2] disease organisms emerging from nature; antibiotic resistance accelerating to the brink of a "post-antibiotic era"; and shrinking global investment in preventive and reactive public health infrastructure.

The deadly, disruptive West African Ebola outbreak of 2014–15 was a serious epidemic—a genuine global public health crisis, made possible by a tepid, late, and poorly resourced response by the international community and major institutions such as the World Health Organization (WHO) (Al Jazeera 2015).[3] By mid-2015, the West African nations of Liberia, Sierra Leone, and Guinea—the hardest hit in the region—had experienced tens of thousands of confirmed, likely, or suspected cases, and over ten thousand probable deaths from Ebola. West Africa's Ebola virus disaster highlighted the world's often fragmented, complacent (at least initially), and inadequately resourced infectious disease infrastructure. The implications were clear: the international public health regime, led by WHO, had suffered significant damage to its credibility, and would require greater funding and at least some restructuring to be able to respond to future such events (Philips and Markham 2014). The answer to the question about whether those lessons would be internalized by U.S. or international policymakers has yet to be determined.

The fall 2014 U.S. Ebola scare,[4] by contrast, was primarily a political and public health communication crisis—though in that regard, also a cautionary warning about what could happen with other diseases posing a far greater risk. As cases turned up in the U.S. for the first time ever during the then-growing epidemic in West Africa, Ebola virus disease dominated much of the U.S. public airways. The good news for Americans is that Ebola demonstrated the American infrastructure to be (on average and despite various stumbles) actually generally well prepared, at least for small-scale U.S. outbreaks of moderate-to-low transmissible viral diseases.[5]

Nevertheless, Ebola and other modern outbreaks illustrate continued problems with U.S. and global infectious disease response capabilities and communications (Ingham and Zingg 2015). Had Ebola landed in the U.S. in a community more off the public health radar (poor, uninsured, illegal immigrants, etc.), there could have been dozens more cases. These issues could and likely will be far more dramatic in the face of future, more transmissible pathogens less deadly than Ebola—including ongoing threats of global pandemic agents such as a novel influenza virus. In such cases, even public health and other *critical infrastructure systems* and *supply chains* in wealthy countries such as the U.S. could be severely challenged by insufficient planning, inadequate medical/public health *surge capacity*,[6] and legal and ethical challenges to response measures. Even far less dramatic pathogens, such as the Zika virus, could have significant impacts on tourism, business, and public reactions that emergency managers and high-level decision makers might need to address.

All of these issues apply to both public- and private-sector preparation/planning and response. This chapter helps frame some of the complexities of the infectious disease dilemma for organizational managers and disaster risk analysts. To do so, it provides some brief examples, in the first section, of actual infectious outbreaks that have occurred in recent years, and might arise again. It also gives guidance, in the second section, on some of what might be done to anticipate, mitigate, and respond adaptively and flexibly to future disease outbreaks, to reduce infectious disease transmission (minimize incidence and effects of illness) among employees, organizational partners, and customers, and to maintain continuity of operations and services.

Some examples of potential disruptive infectious diseases

Among the most disruptive infectious disease disaster scenarios that could arise would be a novel pandemic influenza viral strain that spreads easily, requiring hospitalization for a large percentage of cases, and/or with a relatively high case fatality (mortality) rate. For such a scenario, there might be no widely produced preventive vaccine available for as long as a year or more, and a severe shortage of possible anti-viral medications (e.g., Tamiflu), which in any case might not work on the pandemic strain (Knobler et al. 2005). In such a case, behavioral changes and non-pharmaceutical countermeasures would be all that might be available for a good part of the time. Since pandemics can range from mild (2008) to severe (1918), there is no one-size-fits-all response portfolio, but any flu pandemic could provide as little as a month or less of advanced warning, last more than a year (the great 1918 pandemic lasted at least 18 months), pass in a series of multiple waves lasting for weeks to months at a time, and have widely varying infection and death rates among each wave (Barry 2004, Taubenberger and Morens 2006).

Influenza is arguably the greatest pandemic threat to humanity at the moment, in addition to the potential for a new, non-flu virus emerging from nature—like SARS or

MERS, for instance, but more transmissible. Newly transformed flu strains have the potential to find little immunity among large percentages of the world's population, and thus humanity faces the constant threat of another pandemic. There has been considerable progress on "universal vaccines" for flu, but this promising research is still far from producing usable, effective broad-spectrum vaccines to cover all novel influenzas (Nabel and Fauci 2010, Brown 2015).

Flu is a significant global infectious disease risk for a number of reasons. First among these is the fact that variations of the virus are highly likely to be transmissible by aerosolized means (i.e., through the air) and inhaled, or by contaminated surfaces, and often have the capability to spread fairly easily among either human and/or animal populations. Thus, *unlike* Ebola, most strains of flu that infect humans do not require intimate contact to be passed on, since they can be passed on to secondary victims by either inhalation of airborne particles, or self-infection from contact with contaminated surfaces on which the relatively sturdy virus has been deposited.[7]

In some instances, international travel creates the potential for rapid spread, economic damage, disruption of business and government, and the potential for panic. Ebola, SARS, and MERS have been highly disruptive, and created complicated crises in the societies affected—sparking panic, requiring extensive response measures (including widespread quarantine and isolation,[8] travel screening, and other restrictions), and highlighting the importance of small numbers of "super-spreaders"[9] who single-handedly infected dozens of others.

SARS, caused by the severe acute respiratory syndrome (SARS-CoV) virus, was identified for the first time ever in late 2002 in Asia, as a mysterious and deadly unknown respiratory virus traced to animals sold in markets in China (its final case fatality rate ended up at around 11 percent of clinically identified cases). It was not controlled for almost a year after it first appeared. During its height, SARS disrupted international travel and national economies (Knobler et al. 2004)—and raised fear of a global pandemic.[10] The U.S. managed to avoid any outbreak. One infected SARS super-spreader was a traveler staying in a hotel in Hong Kong, who was linked to more than twenty other cases, including in Canada, Singapore, and Vietnam (McGill 2015, Wong et al. 2015). In fact, one study found that around 75 percent of the Hong Kong and Singapore SARS infections were linked to super-spreader cases. This underscores the critical role of both early diagnosis and proper isolation and infection control measures (Li et al. 2004).

The most recent deadly coronavirus to emerge, Middle Eastern Respiratory Syndrome (MERS-CoV), originated in Saudi Arabia in 2012, and caused great international consternation.[11] MERS was limited to a few hundred cases in the Middle East until May 2015, when a single case imported by a traveler returning from the Middle East started a cascading outbreak in South Korea healthcare facilities. Like SARS, there is, as of this writing, no evidence of widespread airborne (aerosol) transmission or of community-based human-to-human transmission, beyond intimate contact with people or bodily fluids (WHO June 20, July 28, 2015). In South Korea's 2015 outbreak, MERS struck panic into much of the population for several weeks, in part because of a combination of poor preparedness, slow response, government secrecy, and poor public health and medical practices (Jack 2015). One key "super-spreader" waited days in a hospital emergency room (ER), and infected dozens of others before being isolated (Hyun and Park 2015).

While not (yet) in the realm of "extreme risk" at this writing, measles is among the most contagious (transmissible) diseases known to humanity.[12] Measles had made a disturbing comeback in the United States for a handful of years, in part because of cases imported from abroad and because of lagging childhood vaccination rates tied to unwarranted but persistent public fears questioning vaccine safety (Matt 2016).

Another institutionally linked pathogen that occasionally causes significant disruption is bacterial meningitis.[13] Bacterial meningitis has only sporadically broken out in the U.S., but because of the horrific disease symptoms, death rate, and often severe after effects, *Neisseria meningitidis* type C is a much-feared infection (CIDRAP 2012).

For Western nations, in particular, but ultimately for the rest of the world, there is also a growing threat from use of biological weapons. The threat has often been overestimated, especially after the 2001 anthrax attacks, but it is increasing, in large measure because of the pace of advances in biological engineering technology that will increasingly enable easy manipulation and production of biological agents, for nation-states or non-state terrorist groups alike (Suk et al. 2011).

Late in 2015, another long-feared public health threat emerged, when versions of bacterial "superbugs" resistant for the first time to all classes of antibiotics (and therefore untreatable) were discovered in Europe, China, and other locations—and by 2016 had reached the U.S. (Wappes 2016). Each year, the U.S. population experiences around 2 million illnesses and 23 thousand deaths caused by bacteria and fungi resistant to at least some classes of antibiotics—including organisms such as *E. coli, salmonella, Pseudomonas,* and *Klebsiella pneumonia*, among others. Some are already resistant to most antibiotics, and, although not associated with acute pandemics up to now, antibiotic-resistant microbes have become among the greatest threats to public health—a global emergency (Olaitan et al. 2015). Globally, there has been a rise in deadly drug-resistant strains of tuberculosis (TB), and particularly of totally drug-resistant TB, all of which has and will pose a threat within the U.S., given that TB is transmitted through the air from person to person (Parida et al. 2015). The economic cost of these infections is substantial: $20 billion in annual additional healthcare spending (hospital stays with resistant infections are often extremely long), and $35 billion in lost economic/business productivity (CDC 2013).

Practical management response issues for extreme disease outbreak events

The world faces a growing risk from infectious disease outbreaks, for many reasons. "Exotic" infectious diseases can hitch a ride anywhere and arrive in any community or region in the world. Those that are transmissible from person to person could rapidly gain a foothold in even the wealthiest nations, including the U.S., particularly if public health has let down its guard.

Important risk variables for infectious diseases, and for overall future U.S. and global risk, include vulnerabilities related to global trade and travel, the need for better investment in surveillance and WHO "rapid strike" capabilities, and better medical and public health infrastructure in developing countries, where outbreaks are likely to arise.

Much has been written on this subject (especially post-9/11, post-anthrax), and much non-terrorism related infectious disease planning focuses on influenza pandemic planning, most of it aimed at governments (WHO 2013), but some targeted at businesses (see for instance OSHA 2007, San Francisco Department of Public Health 2008). Using case studies and best practices, we here address a number of issues, including: practical medical and non-medical countermeasures; the interplay of government, private sector, and citizenry; business continuity and supply-chain issues; legal issues that might arise during outbreaks; and other important aspects of infectious disease disaster management.

Be prepared to follow the recommendations of local and state public health

The U.S. Constitution gives primacy to states in matters of public health, except in the most extreme national emergencies. Public health laws vary widely around the country, and even within some states, but emergency managers still need to comply with local and state public health recommendations and orders.

Public health advice during a major outbreak of a person-to-person transmissible/contagious outbreak (such as severe pandemic flu) would likely include recommendations or orders intended to stop or slow down the spread, such as:

- social distancing (avoiding contact with potentially sick people, postponing public gatherings, closing schools and non-essential businesses);
- wearing surgical masks or other face masks in public;
- voluntary or involuntary quarantine of suspected exposed/infected persons (for an examination of the legal issues involved in this, see for instance Cole 2014 and Parmet 2007);
- isolation of the ill (for days or weeks) until they are no longer a risk;
- good personal hygiene (hand washing, proper cough etiquette, etc.).

Businesses and other organizations should be aware that they could risk legal liability from disregarding orders or recommendations related to their business conduct, both because of possible violations of public health orders or existing laws, and because they might open themselves to lawsuits from injured employees, customers, or business associates who might be exposed to infectious disease in a workplace. That said, public health officials usually issue recommendations rather than edicts. Managers should also be prepared for possible disruptions of supply chains, personnel absenteeism due to illness and perceptions of disease threats (beyond normal cold and flu season), and potential business continuity problems (as discussed below in more detail).

Also important is the need to be prepared for plans to fail. Too many planners assume away uncertainty, and cultivate hubris. As Clarke (1999) notes, this often leads to "fantasy plans," based on unrealistic assumptions of certainty and inadequate information. These plans can be worse than nothing, since they convey a false sense of preparedness.

Infectious disease disasters are usually either unknown or unfamiliar for most civilian non-experts who would be managing such an event, and because of this the need for contingency planning is even greater. Unlike "fast" disasters like earthquakes, tornadoes, hurricanes, or most terrorist attacks (which are usually over in a short period of time), infectious disease outbreaks can last for days, weeks, months, or even years. Since most infectious diseases are contingent on a mix of human behavior and a disease pathogen's particular "personality" (transmissibility, lethality, treatment, or prevention issues, etc.), predicting the course of outbreaks/epidemics can be difficult. It is therefore important to prepare for a wide range of scenarios, from mild to severe, and to be aware that every outbreak has a different dynamic.

Paid sick days are proven to save money and time

The U.S. Bureau of Labor calculates that 39 percent (41 million people) of all American workers do not have paid sick leave (Perez 2015). Not only during declared epidemics but also in everyday operations, research has demonstrated that allowing employees paid sick days both helps their productivity and prevents others from getting sick in the workplace. Various academic studies have demonstrated that "presenteeism" (vs. absenteeism) costs are "higher than the combined costs of medical care, prescription drugs and absenteeism," leading to $150 billion to $250 billion a year in business losses—around 60 percent of total productivity losses (Templeton 2015).

In the early days of a major infectious disease disaster event, where staying home sick could reduce the risk of other key employees being sickened in the workplace, lack of paid sick leave could cause a major blow to business continuity. As discussed below, dealing with employee absenteeism is a central element of maintaining key operations.

Prevention, early detection, and early response can be critical

Prevention, and/or early response, is critical for infectious disease outbreaks. Emergency managers in the workplace can be instrumental in urging sick employees to either stay home, or visit available medical facilities, instead of going to work. In some cases, they can even help facilitate adult vaccination.

As noted by Hagerman et al. (Chapter 5 this volume), prevention and early detection and response are far more important for some diseases than others. Ebola, for instance, is only moderately contagious, but highly deadly.[14] Measles, by contrast, is highly contagious, but not as deadly (though measles remains a major cause of death worldwide for children under five). Both therefore require rapid public health response and complicated institutional decisions, for different reasons.

Personnel and staffing issues for business continuity

Numerous issues can affect the probability of maintaining essential services and personnel, including perceptions of workplace safety, actual workplace risk mitigation (including procedures, safety equipment, sick leave, and work-at-home policies), isolation and quarantine, and other factors. If the workplace is considered unsafe in a public health crisis, critical staff may opt to stay home.

In a severe pandemic flu scenario, during the height of a major wave of disease, businesses and other organizations can expect a significant percent of their workforce to be absent. Estimates range from 20 to 50 percent at any given time due to personal illness (which could be severe enough to require weeks of recuperation for survivors). Others will need to stay home with sick children or other relatives, or will simply fear exposure. Those with children in school or childcare settings may be particularly susceptible. Death rates from flu could range from a fraction of one percent to high single digits.

In planning for such disasters, it is important to assure people that their families will be protected. Depending on the scope of any particular infectious disease event, providing masks/respirators, hand sanitizer, and even food assistance if grocery store shelves are bare can increase willingness to work.[15] Similarly, to minimize workplace/public exposure, policies and planning for having white-collar employees (especially) work remotely (*telecommute*) from home should be addressed before a crisis. As long as the internet and communications systems are still functioning properly, such plans can maintain many operations.

Decisions about paying staff who are or may be sick and self-isolating or voluntarily quarantining themselves should also be considered ahead of time. Poorly designed, unfair, or randomly enforced policies in an emergency can damage morale and increase the possibility of absenteeism.

Basic infection control/hygiene measures

Good infection control/hygiene measures and procedures can help reduce organizational risks from transmissible infectious outbreaks for essential personnel.

In the 2014 U.S. Ebola scare, the first case identified, Eric Thomas Duncan, illustrated the dangers of lax training and protocols for personal protective equipment (PPE) and hygiene. The failure to protect clinical personnel contributed to the infection of two nurses treating that patient (Gambino 2014).

The importance of stringent infection control and prevention practices cannot be overestimated in a clinical healthcare setting. These can also, to a far lesser extent, be employed by businesses and other organizations in the event of an infectious disease disaster, such as a pandemic flu, to help maintain continuity of operations and protect essential personnel who could be exposed in a work/organizational environment. While untrained employees are less likely to follow such measures or to use protective equipment properly (or at all), such measures could reduce the overall infection risk, particularly if employees are given assistance in competently fitting and using items like masks (which cannot be used effectively with beards or even stubble, for instance).

To prepare for possible worst-case outbreaks, organizations should consider stockpiling various supplies, including hand sanitizer, single-use low filtration medical masks and/or higher filtration N95 respirator masks,[16] and even eye protection (goggles), disposable gowns, and gloves. Strict hand washing before and after contact with people or objects has been shown to be significant in reducing infection. The objective of such measures is to avoid contact with the infectious agent or body fluid that may contain the infectious agent by creating a barrier between the worker and the infectious material. Gloves protect the hands, gowns or aprons protect the skin and/or clothing, masks and respirators protect the mouth and nose, goggles protect the eyes, and face shields protect the entire face (CDC 2004).

For businesses and non-medical/clinical organizations, the level of equipment needed will vary depending on size, levels of person-to-person contact, and availability of inside expertise and outside assistance. For critical infrastructure operations (electric power, water treatment, public safety, etc.), these measures may be unavoidable if essential employees are expected to report for duty. Critical supplies, such as PPE and even food supplies, could be unreliable in a larger-scale outbreak, particularly in an era of just-in-time supply chains. Ideally, a reserve "surge" capacity of all critical items should be maintained by both smaller and larger organizations anxious to maintain continuity of operations.

Like any planning and preparation, however, it's not just the equipment and supplies that are important, it's the ability of people in the organization to use them. In many instances this involves training and even drilling of emergency plans. This is especially true for use of PPE and hygiene procedures.

Supply-chain issues

Significant absenteeism in severe outbreaks can challenge horizontal and vertical supply chains. If there are too few people available to produce, transport, and retail products and services, this will stress both upstream and downstream parts of the system. In a just-in-time world, supply chain management in a disaster will often be affected by the weakest link, so ideally planning should include collaboration or consultation with an organization's suppliers. Otherwise, contractors, vendors, truckers, and other private operations may be highly unreliable at the height of a pandemic.

Studies indicate that in all but the most severe infectious disease disasters, basic government and private essential municipal services (government, electric, public safety, water, sewage, trash) will continue to operate, even if stressed, given critical infrastructure planning that has already occurred (especially since 9/11). However, in a major disaster, organizations may be on their own at times, given excess demand and similar stress on nearby mutual aid, state, and federal services. In other words, organizations and their personnel may not be able to rely on "the cavalry" to arrive for some time.

There will be economic and institutional-trust damage

In addition to direct economic costs from care for the ill and dead, minimizing the indirect economic costs is often even more important, particularly due to lost productivity and disruption of key infrastructure and supply chains. These disruptions from infectious disease outbreaks can be the largest societal costs of an epidemic.[17] Border shutdowns and travel restrictions, quarantine of the healthy, isolation of the sick, and a climate of fear that confounds cooperation and efficiency during an outbreak can create rumors and resentments, fueled by social media, with the potential to severely damage institutional reputations/brands for government and private-sector organizations alike.

Public perception in the internet age can be volatile. Despite what experts may advise, rumor mills can quickly spread misinformation or outright lies.

Among the lessons learned for maintaining trust include:

- Don't lie! Institutional credibility is gold and difficult to recoup when lost; it's better to say you don't know but are taking steps to find the answer.
- Government and private-sector secrecy, poor messaging, poor infection control, and a lack of preparedness increase the potential for fear and panic, among both organizational personnel and the public, as happened at various points during the 2014 U.S. Ebola scare.
- Sometimes reassurance/confidence-building measures, even if not seen to be effective by experts, can help maintain public or civilian employee cooperation. Ohio and Texas schools during the U.S. Ebola outbreak fumigated their facilities because of "irrational" public fears, even though there was essentially no risk of infection. Such measures can allow normal functioning to return in situations where people can vote with their feet about whether to report to work.

Conclusion

Emerging infectious disease outbreaks are increasing in frequency, particularly as climate change disrupts the global environment. Not all outbreaks are equal, and response is subject to surge, where a large outbreak can overwhelm even the best systems. However, even a relatively small infectious disease outbreak can overwhelm public health and medical surge capacities, depending on a host of variables, including the nature of the pathogen, its environment, and any viable prevention or treatment options; public reaction; and institutional response efficiency and priorities. This has been amply demonstrated in events such as the 2011 anthrax attacks (which cost billions of dollars in response, for only five deaths and less than two dozen cases) and the 2014 U.S. Ebola scare.

The 2014–15 West African Ebola epidemic was a shot across the bow of international public health. Rapid response capabilities, akin to a public health wildfire response team, must be in place to minimize the impact of outbreaks before they blossom into catastrophes. As critical as adaptive preparation and response are, preventive mitigation can be far more cost effective (Pike et al. 2014). Public health surveillance at the source of potential pandemics was once robust but has become a neglected public good in an era of budget cuts and austerity.

Domestic U.S. infectious disease disaster risk analysis and policy must account for the unique, complex nature of pathogenic organisms, and the sometimes unpredictable realities of public and political reaction. There is a rising risk from outbreaks of novel diseases such as SARS, MERS, Zika virus, and the ongoing, omnipresent threat from novel strains of influenza, and even biological terrorism. All of this involves reinvestment in cost-effective infrastructure and expertise for both prevention and emergency response.

Risk perception influences people's compliance with public health recommendations and their demands for response from government officials and the private sector. In an age of 24-hour news cycles, profit-driven infotainment, easy political opportunism, and lightning-fast internet, what's efficient, sound, or even legal from a public-health expert perspective may not be what politicians and those in charge of government or private organizations end up doing. By being prepared for a range of possibilities, risk managers and crisis responders can minimize the disruption from even low-probability, high-impact events such as major infectious outbreaks—a task that is crucial in an era when such diseases can be around the world in less than 48 hours.

Notes

1 Bill Gates, "The Rational Optimist: How Prosperity Evolves." Gatenotes blog (November 10, 2010) http://www.gatesnotes.com/Books/Africa-Needs-Aid-Not-Flawed-Theories

2 For the purposes of this discussion, a *pathogen* is simply a disease-causing agent or microorganism, usually associated with viruses, bacteria, protozoa, fungi, prions, or other sources. An *epidemic* is a disease outbreak (occurrence of disease cases) that is larger than what is normally seen in a particular area, and a *pandemic* is a larger geographic epidemic, occurring across a large region or even around the world.

3 One notable exception being non-governmental organizations that had already been operating in the region, in particular Médecins Sans Frontières (MSF—a.k.a., Doctors Without Borders), which provided a critical if lonely early role that likely prevented a far worse outcome (MSF 2015).

4 In the U.S. there was one death from an imported case (Liberian national Eric Duncan), two nurses infected from that case, and a few other diagnosed domestic cases among healthcare workers infected in and returned from West Africa, all of whom recovered.

5 In as much as most U.S. public health infrastructure and response is by Constitutional mandate regulated at the state and even local levels, there is a great deal of variation in those capabilities across the country.

6 The ability of any system to adequately absorb a larger than normal "surge" of demand for services, infrastructure, etc.—from expert personnel to hospital beds and medicine.

7 Depending on environmental conditions on said surfaces, the flu virus can live for hours or even days and still remain infectious.

8 Quarantine is separating people who are suspected of exposure/infection, but not symptomatic, from the rest of the population; isolation is doing so to those who are already sick, and can be either voluntary or involuntary (Rothstein 2004).

9 A "super-spreader" for a given disease either transmits higher amounts of pathogen than other patients, or for environmental or other reasons infects more people than average sick individuals do, usually in a context with either inadequate or non-existent infection control procedures. These small numbers of outlier cases are in contrast to the concept of R_0, the average reproductive rate at which a transmissible disease is passed on to infect other generations. Super-spreaders have been implicated in a majority of severe outbreaks (Stein 2011).

10 Though not a pandemic, per se, major 2003 SARS outbreaks occurred in Asia (particularly China, where it started) and Canada, and lesser outbreaks in Australia, Europe, Africa, and South America, and particularly the cities of Hong Kong, Beijing, Singapore, and Toronto, Canada.

11 Since first detection in Saudi Arabia, by late August 2015, MERS-CoV had killed 492 out of 1,154 confirmed cases (over 40 percent case fatality rate) in Saudi Arabia, according to the Ministry of Health (Schnirring 2015).

12 The R_0 of measles ranges as high as 12–18 cases (in unvaccinated or less-immune populations) infected by a single sick individual, absent proper infection control measures (Orenstein and Gay 2004), unlike primarily droplet transmission pathogens (such as Ebola, SARS, MERS).

13 Viral meningitis is far less dangerous than its bacterial disease "cousin."

14 Even if it had gained a foothold "off the radar" (for instance, in undocumented resident alien communities), Ebola would likely not have been widely disseminated, since transmission requires relatively intimate, close contact, and is not significantly contagious until victims experience comparatively severe disease symptoms.

15 During the 2003 SARS outbreak, a substantial number of healthcare workers treating patients in Toronto, Canada, became sick, despite standard precautions. Even worse, some brought SARS home to their families, striking fear into the healthcare community. Anecdotal reports suggested that, by the time that outbreak ended, many were at the point of not showing up for work because of their fear (author's anonymous discussion with a Toronto ER physician in early 2004).

16 See, for instance, NIOSH-approved N95 Particulate Filtering Facepiece Respirators http://www.cdc.gov/niosh/npptl/topics/respirators/disp_part/n95list1.html

17 Analogous to the aviation industry's losses (for years) after the 9/11 terrorism attacks, resulting from a 25 percent drop in people flying. These fear-based secondary losses ultimately exceeded the cost of the attacks themselves.

References

Al Jazeera (2015). "Doctors Without Borders slams global response to Ebola," *Al Jazeera* (March 22, 2015). http://www.aljazeera.com/news/2015/03/doctors-borders-slams-global-response-ebola-150322233606792.html

Barry, John M. (2004). *The Great Influenza*. Penguin Books, New York.

Brown, Rachael (2015). "Universal flu vaccine on the horizon after successful animal tests." *The World Today* (August 25, 2015). http://www.abc.net.au/worldtoday/content/2015/s4299672.htm

Centers for Disease Control and Prevention (CDC) (2004). "Guidance for the selection and use of personal protective equipment (PPE) in healthcare settings." https://www.cdc.gov/HAI/pdfs/ppe/PPEslides6-29-04.pdf

Centers for Disease Control and Prevention (CDC) (2014). "Influenza viruses: how the flu virus can change: 'drift' and 'shift.'" http://www.cdc.gov/flu/about/viruses/change.htm

Centers for Disease Control and Prevention (CDC) (2013) "Antibiotic resistance threats in the united states, 2013." Centers for Disease Control and Prevention, U.S. Department of Health and Human Services: Atlanta, GA. https://www.cdc.gov/drugresistance/threat-report-2013/index.html

CIDRAP (2012). "Meningitis outbreak toll rises as complications surface." CIDRAP (November 5,2012).http://www.cidrap.umn.edu/news-perspective/2012/11/meningitis-outbreak-toll-rises-complications-surface

Clarke, Lee (1999). *Mission Improbable: Using Fantasy Documents to Tame Disaster*. University of Chicago Press, Chicago, IL.

Cole, Jared P. (2014). *Federal and State Quarantine and Isolation Authority*. RL33201. (October 9, 2104). Congressional Research Service, Washington, DC.

Gambino, Lauren (2014). "Thomas Eric Duncan's family settles with Dallas hospital over Ebola treatment," *The Guardian* (November 12, 2014). https://www.theguardian.com/world/2014/nov/12/thomas-eric-duncan-family-settlement-dallas-hospital

Hyun, Myung Han, and Yoonjee Park (2015). "Re: Why the panic? South Korea's MERS response questioned." *BMJ*, 350:h3403. http://www.bmj.com/content/350/bmj.h3403/rr

Ingham, Richard, and Elisabeth Zingg (2015). "The Ebola Epidemic Is Causing People to Lose Faith in the Medical Order." *Business Insider* (October 26, 2014). http://www.businessinsider.com/ebola-epidemic-lose-faith-medical-order-2014-10

Jack, Andrew (2015). "Why the panic? South Korea's MERS response questioned." *BMJ*, 350:h3403. https://doi.org/10.1136/bmj.h3403

Knobler, Stacey, A. Mahmoud, S. Lemon, A. Mack, L. Sivitz, and K. Oberholtzer (eds) (2004). *Learning From SARS: Preparing for the Next Disease Outbreak: Workshop Summary*. U.S. Institute of Medicine. National Academies Press, Washington, DC.

Knobler S, A. Mack, A. Mahmoud, S. Lemon (eds) (2005). *The Story of Influenza: The Threat of Pandemic Influenza: Are We Ready?* Workshop Summary. The National Academies Press, Washington, D.C.

Li, Y., I.T. Yu, P. Xu, J.H. Lee, T.W. Wong, P.L. Ooi, et al. (2004). "Predicting super spreading events during the 2003 SARS epidemics in Hong Kong and Singapore." *American Journal Epidemiology*, 160, 719–728.

Matt, Liana (2016). "Study relates vaccine refusal to rise in measles, pertussis." CIDRAP News (March21,2016).http://www.cidrap.umn.edu/news-perspective/2016/03/study-relates-vaccine-refusal-rise-measles-pertussis

Médecins Sans Frontières (MSF) (2015). "Ebola: Pushed to the limit and beyond: A critical analysis of the global Ebola response one year into the deadliest outbreak in history." Médecins Sans Frontières (March 23, 2015). http://www.msf.org/en/article/ebola-pushed-limit-and-beyond

McGill, Natalie (2015). "Public health looks to lessons from past in MERS outbreak: Experience with SARS offers insights." *The Nation's Health,* 45(6), 1–12. http://thenationshealth.aphapublications.org/content/45/6/1.4.long

Nabel, Gary J., and Anthony S. Fauci (2010). "Induction of unnatural immunity: prospects for a broadly protective universal influenza vaccine." *Nature Medicine*, 16(12), 1389–1391.

Olaitan, Abiola Olumuyiwa, Serge Morand, and Jean-Marc Rolain (2015). "Emergence of colistin-resistant bacteria in humans without colistin usage: a new worry and cause for vigilance." *International Journal of Antimicrobial Agents*, 47(1), 1–3.

Occupational Safety and Health Administration (2007). *Guidance on preparing workplaces for an influenza pandemic.* Occupational Safety and Health Administration , Washington, DC. http://digitalcommons.ilr.cornell.edu/key_workplace/617

Orenstein, W.A., and N.J. Gay (2004). "The theory of measles elimination: implications for the design of elimination strategies." *Journal of Infectious Diseases*, *189*(Supplement 1), S27–S35.

Parida, S.K., R. Axelsson-Robertson, M.V. Rao, N. Singh, I. Master, A. Lutckii, ... and M. Maeurer (2015). "Totally drug-resistant tuberculosis and adjunct therapies." *Journal of Internal Medicine*, 277(4), 388–405. http://onlinelibrary.wiley.com/doi/10.1111/joim.12264/full

Philips, Mit, and Áine Markham (2014). "Ebola: a failure of international collective action." *The Lancet*, 384(9949), 1181.

Pike, Jamison, Tiffany Bogich, Sarah Elwood, David C. Finoff, and Peter Daszak (2014) "Economic optimization of a global strategy to address the pandemic threat." *Proceedings of the National Academy of Sciences*, 111(52), 18519–18523

Parmet, W.E. (2007). "Legal power and legal rights: Isolation and quarantine in the case of drug-resistant tuberculosis." *New England Journal of Medicine*, *357*(5), 433–435.

Perez, Tom (2015). "Sick and tired: new BLS data highlights the need for paid leave." U.S. Department of Labor (July 27, 2015). https://blog.dol.gov/2015/07/27/sick-and-tired-new-bls-data-highlights-the-need-for-paid-leave/

Rothstein, Mark A. (2004). "Are traditional public health strategies consistent with contemporary American values." *Temple Law Review*, 77, 175–192.

Rothstein, M.A., M.G. Alcalde, N.R Elster, M.A Majumder, L.I. Palmer, ... and R.E. Hoffman (2003). *Quarantine and Isolation: Lessons Learned from SARS*. University of Louisville School of Medicine, Institute for Bioethics, Health Policy and Law.

San Francisco Department of Public Health (2008). *Pandemic Influenza Business Continuity Guide & Template For San Francisco Businesses*. (July 2008). *www.sfcdcp.org/document.html?id=260*

Schnirring, Lisa (2015). "MERS Cases from Riyadh still surging." CIDRAP (August 24, 2015). http://www.cidrap.umn.edu/news-perspective/2015/08/mers-cases-riyadh-still-surging

Stein, Richard A. (2011). "Super-spreaders in infectious diseases." *International Journal of Infectious Diseases*, 15(8), e510–e513. http://www.sciencedirect.com/science/article/pii/S1201971211000245

Suk, J. E., A. Zmorzynska, I. Hunger, W. Biederbick, J. Sasse, H. Maidhof, and J.C. Semenza (2011). "Dual-use research and technological diffusion: reconsidering the bioterrorism threat spectrum." *PLoS Pathogens*, 7(1), e1001253. https://doi.org/10.1371/journal.ppat.1001253

Taubenberger, Jeffery K., and David M. Morens (2006). "1918 influenza: the mother of all pandemics." *Emerging Infectious Diseases*, 12(1), 15–22. doi:10.3201/eid1201.050979

Templeton, David (2015). "Is coming to work sick really a good idea?" *Pittsburgh Post-Gazette* (December 15, 2015). http://www.post-gazette.com/news/health/2015/12/15/Presenteeism-more-costly-than-absenteeism/stories/201511040201

Wappes, John (2016). "Highly resistant MCR-1 'superbug' found in U.S. for first time." CIDRAP (26 May 2016). http://www.cidrap.umn.edu/news-perspective/2016/05/highly-resistant-mcr-1-superbug-found-us-first-time

Wong, Gary, Wenjun Liu, Yingxia Liu, Boping Zhou, Yuhai Bi, George F. Gao (2015). "MERS, SARS, and Ebola: the role of super-spreaders in infectious disease." *Cell Host and Microbe*, 18(4), 398–401. http://dx.doi.org/10.1016/j.chom.2015.09.013

World Health Organization (2013). *Pandemic influenza risk management: WHO interim guidance*. World Health Organization, Geneva, Switzerland.

World Health Organization (June 20, 2015). "Middle East respiratory syndrome coronavirus (MERS-CoV)–Thailand." http://who.int/csr/don/20-june-2015-mers-thailand/en/

World Health Organization (July 28, 2015). "Intensified public health measures help control MERS-CoV outbreak in the Republic of Korea." http://www.wpro.who.int/mediacentre/releases/2015/20150728/en/

11 Global catastrophes

The most extreme risks

Seth D. Baum and Anthony M. Barrett

Introduction

The most extreme type of risk is the risk of a global catastrophe causing permanent worldwide destruction to human civilization. In the most extreme cases, human extinction could occur. Global catastrophic risk (GCR) is thus risk of events of the highest magnitude of consequences, and the risks merit serious attention even if the probabilities of such events are low. Indeed, a growing chorus of scholars rates GCR reduction as among the most important priorities for society today. Unfortunately, many analysts also estimate frighteningly high probabilities of global catastrophe, with one even stating, "I think the odds are no better than fifty-fifty that our present civilization on Earth will survive to the end of the present century" (Rees 2003:8).

Regardless of what the probabilities are, it is clear that humanity today faces a variety of serious GCRs. To an extent, humanity always has faced GCRs, in the form of supervolcano eruptions, impacts from large asteroids and comets, and remnants of stellar explosions. Events like these have contributed to several mass extinction events across Earth's history. The Toba volcano eruption about 70,000 years ago may have come close to bringing the human species to a premature end. However, scholars of GCR generally believe that today's greatest risks derive from human activity. These GCRs include war with biological or nuclear weapons, extreme climate change and other environmental threats, and misuse or accidents involving powerful emerging technologies like artificial intelligence (AI) and synthetic biology. These GCRs threaten far greater destruction than was seen in the World Wars, the 1918 flu, the Black Death plague, or other major catastrophes of recent memory.

The high stakes and urgent threats of GCR demand careful analysis of the risks and the opportunities for addressing them. However, several factors make GCR difficult to analyze. One factor is the unprecedented nature of global catastrophes. Many of the catastrophes have never occurred in any form, and of course no previous global catastrophe has ever destroyed modern human civilization. The lack of precedent means that analysts cannot rely on historical data as much as they can for smaller, more frequent events. Another factor is the complexity of GCRs, involving global economic, political, and industrial systems, which present difficult analytical decisions about which details to include. Finally, the high stakes of GCR pose difficult dilemmas about the extent to which GCR reduction should be prioritized relative to other issues.

In this paper we present an overview of contemporary GCR scholarship and related issues for risk analysis and risk management. We focus less on the risks themselves, each of which merits its own dedicated treatment. Other references are recommended for the risks, perhaps the best of which are the relevant chapters of Bostrom and Ćirković (2008). Instead, our focus here is on overarching themes of importance to the breadth of the GCRs. The following section defines GCR in more detail and explains why many researchers consider it to be so important. Next, some of the analytical challenges that GCR poses and the techniques that have been developed to meet these challenges are explained. There follows a discussion of some dilemmas that arise when GCR reduction would require great sacrifice or would interfere with each other. Finally, conclusions are drawn.

What is GCR and why is it important?

Taken literally, a global catastrophe can be any event that is in some way catastrophic across the globe. This suggests a rather low threshold for what counts as a global catastrophe. An event causing just one death on each continent (say, from a jet-setting assassin) could rate as a global catastrophe, because surely these deaths would be catastrophic for the deceased and their loved ones. However, in common usage, a global catastrophe would be catastrophic for a significant portion of the globe. Minimum thresholds have variously been set around ten thousand to ten million deaths or $10 billion to $10 trillion in damages (Bostrom and Ćirković 2008), or death of one quarter of the human population (Atkinson 1999; Hempsell 2004). Others have emphasized catastrophes that cause long-term declines in the trajectory of human civilization (Beckstead 2013) that human civilization does not recover from (Maher and Baum 2013), that drastically reduce humanity's potential for future achievements (Bostrom 2002a, using the term "existential risk"), or that result in human extinction (Matheny 2007; Posner 2004).

A common theme across all these treatments of GCR is that some catastrophes are vastly more important than others. Carl Sagan was perhaps the first to recognize this, in his commentary on nuclear winter (Sagan 1983). Without nuclear winter, a global nuclear war might kill several hundred million people. This is obviously a major catastrophe, but humanity would presumably carry on. However, with nuclear winter, per Sagan, humanity could go extinct. The loss would be not just an additional four billion or so deaths, but the loss of all future generations. To paraphrase Sagan, the loss would be *billions and billions* of lives, or even more. Sagan estimated 500 trillion lives, assuming humanity would continue for ten million more years, which he cited as typical for a successful species.

Sagan's 500 trillion number may even be an underestimate. The analysis here takes an adventurous turn, hinging on the evolution of the human species and the long-term fate of the universe. On these long time scales, the descendants of contemporary humans may no longer be recognizably "human." The issue then is whether the descendants are still worth caring about, whatever they are. If they are, then it begs the question of how many of them there will be. Barring major global catastrophe, Earth will remain habitable for about one billion more years until the Sun gets too warm and large. The rest of the Solar System, Milky Way galaxy, universe, and (if it exists) the multiverse will remain habitable for a lot longer than that (Adams and Laughlin 1997), should our descendants gain the capacity to migrate there. An open question in astronomy is whether it is possible for the descendants of humanity to continue living for an infinite length of time or instead merely an astronomically large but finite length of time (see e.g. Ćirković 2002; Kaku 2005). Either way, the stakes with global catastrophes could be much larger than the loss of 500 trillion lives.

Debates about the infinite versus the merely astronomical are of theoretical interest (Bossert et al. 2007; Ng 1991), but they have limited practical significance. This can be seen when evaluating GCRs from a standard risk-equals-probability-times-magnitude framework. Using Sagan's 500 trillion lives estimate, it follows that reducing the probability of global catastrophe by a mere one-in-500-trillion chance is of the same significance as saving one human life. Phrased differently, society should try 500 trillion times harder to prevent a global catastrophe than it should to save a person's life. Or, preventing one million deaths is equivalent to a one-in-500-million reduction in the probability of global catastrophe. This suggests society should make extremely large investment in GCR reduction, at the expense of virtually all other objectives.

Judge and legal scholar Richard Posner made a similar point in monetary terms (Posner 2004). Posner used \$50,000 as the value of a statistical human life (VSL) and 12 billion humans as the total loss of life (double the 2004 world population); he describes both figures as significant underestimates. Multiplying them gives \$600 trillion as an underestimate of the value of preventing global catastrophe. For comparison, the United States government typically uses a VSL of around \$1–10 million (Robinson 2007). Multiplying a \$10 million VSL with 500 trillion lives gives $\$5\times10^{21}$ as the value of preventing global catastrophe. But even using "just" \$600 trillion, society should be willing to spend at least that much to prevent a global catastrophe, which converts to being willing to spend at least \$1 million for a one-in-500-million reduction in the probability of global catastrophe. Thus, while reasonable disagreement exists on how large a VSL to use and how much to count future generations, even low-end positions suggest vast resource allocations should be redirected to reducing GCR. This conclusion is only strengthened when considering the astronomical size of the stakes, but the same point holds either way. The bottom line is that, as long as something along the lines of the standard risk-equals-probability-times-magnitude framework is being used, then even tiny GCR reductions merit significant effort. This point holds especially strongly for risks of catastrophes that would cause permanent harm to global human civilization.

The discussion thus far has assumed that all human lives are valued equally. This assumption is not universally held. People often value some people more than others, favoring themselves, their family and friends, their compatriots, their generation, or others whom they identify with. Great debates rage on across moral philosophy, economics, and other fields about how much people should value others who are distant in space, time, or social relation, as well as the unborn members of future generations. This debate is crucial for all valuations of risk, including GCR. Indeed, if each of us only cares about our immediate selves, then global catastrophes may not be especially important, and we probably have better things to do with our time than worry about them.

While everyone has the right to their own views and feelings, we find that the strongest arguments are for the widely held position that all human lives should be valued equally. This position is succinctly stated in the United States Declaration of Independence, updated in the 1848 Declaration of Sentiments: "We hold these truths to be self-evident: that all men and women are created equal." Philosophers speak of an agent-neutral, objective "view from nowhere" (Nagel 1986) or a "veil of ignorance" (Rawls 1971) in which each person considers what is best for society irrespective of which member of society they happen to be. Such a perspective suggests valuing everyone equally, regardless of who they are or where or when they live. This in turn suggests a very high value for reducing GCR, or a high degree of priority for GCR reduction efforts.

Challenges to analyzing GCR

Given the goal of reducing GCR, one must know what the risks are and how they can be reduced. This requires diving into the details of the risks themselves—details that we largely skip in this paper—but it also requires attention to a few analytical challenges.

The first challenge is the largely unprecedented nature of global catastrophes. Simply put, modern human civilization has never before ended. There have been several recent global catastrophes of some significance, the World Wars and the 1918 flu among them, but these clearly did not knock civilization out. Earlier catastrophes, including the prehistoric mass extinction events, the Toba volcanic eruption, and even the Black Death plague, all occurred before modern civilization existed. The GCR analyst is thus left to study risks of events that are in some way untested or unproven. But the lack of historical precedent does not necessarily imply a lack of ongoing risk. Indeed, the biggest mistake of naïve GCR analysis is to posit that, because no global catastrophe has previously occurred, therefore none will occur. This mistake comes in at least three forms.

The first and most obviously false form is to claim that unprecedented events never occur. In our world of social and technological innovation, it is easy to see that this claim is false. But accounting for it in risk analysis still requires some care. One approach is to use what is known in probability theory as zero-failure data (Bailey 1997; Hanley 1983; Quigley and Revie 2011). Suppose that no catastrophe has occurred over n prior time periods—for example, there has been no nuclear war in the 65 years since two countries have had nuclear weapons. (The second country to build nuclear weapons was the Soviet Union, in 1949.) It can thus be said that there have been zero failures of nuclear deterrence in 65 cases. An approximate upper bound can then be estimated for the probability p of nuclear deterrence failure, i.e. the probability of nuclear war, occurring within an upcoming year. Specifically, p lies within the interval $[0, u]$ with $(1 - \alpha)$ confidence, where $u = 1 - \alpha^{(1/n)}$ gives the upper limit of the confidence interval. Thus for 95% confidence ($\alpha = 0.05$), $u = 1 - 0.05^{(1/65)} = 0.05$, meaning that there is a 95% chance that the probability of nuclear war within an upcoming year is somewhere between 0 and 0.05. Note that this calculation assumes (perhaps erroneously) that the 65 non-failures are independent random trials and that p is approximately constant over time, but it nonetheless provides a starting point for estimating the probability of unprecedented events. Barrett et al. (2013) use a similar approach as part of a validation check of a broader risk analysis of U.S.–Russia nuclear war.

The second form of the mistake is to posit that the ongoing existence of human civilization proves that global catastrophes will not occur. It is true that civilization's continued existence despite some past threats should provide some comfort, but it should only provide *some* comfort. Consider this: if a global catastrophe had previously occurred, nobody would still be around to ponder the matter (at least for catastrophes causing human extinction). The fact of being able to observe one's continued survival is contingent upon having survived. While it is easy to see that this is a mistake, it is harder to correct for it. Again, it requires careful application of probability theory, correcting for what is known as an observation selection effect (Bostrom 2002b; Ćirković et al. 2010). The basic idea is to build the existence of the observer into probability estimates for catastrophes that would eliminate future observers. The result is probability estimates unbiased by the observer's existence, with global catastrophe probability estimates typically revised upwards.

The third form of the mistake is to posit that, because humanity has survived previous catastrophes, or risks of catastrophes, therefore it will survive future ones. This mistake is especially pervasive in discussions of nuclear war. People sometimes observe that no nuclear war has ever occurred and cite this as evidence to conclude that therefore nuclear deterrence and the fear of mutually assured destruction will indefinitely continue to keep the world safe (for discussion see Sagan and Waltz 2013). But there have been several near misses, from the 1962 Cuban missile crisis to the 1995 Norwegian rocket incident, and there is no guarantee that nuclear war will be avoided into the distant future. Similarly, just because no pandemic has ever killed the majority of people (Black Death killed about 22 percent), or just because early predictions about the rise of artificial intelligence proved false (they expected human-level AI within decades that have long since come and gone; see Crevier 1993; McCorduck 2004), it does not necessarily follow that no pandemics would be so lethal, or that AI cannot reach the lofty heights of the early predictions.

Careful risk analysis can correct for the third form by looking at the full sequences of events that would lead to particular global catastrophes. For example, nuclear weapons in the United States are launched following a sequence of decisions by increasingly high ranking officials, ultimately including the President. This decision sequence can be built into a risk model, with model parameters estimated from historical data on how often each step in the decision sequence has been reached (Barrett et al. 2013). The more often near misses have occurred, and the nearer the misses were, the higher the probability of an eventual "hit" in the form of a nuclear war. The same analytic structure can be applied to other GCRs.

But for many aspects of GCRs, as with other low-probability risks, there is not enough historical or other empirical data to fully characterize the risk. A good example of this is the risk from AI. The concern is that AI with human-level or super-human intelligence could outsmart humanity, assume control of the planet, and inadvertently cause global catastrophe while pursuing whatever objectives it was initially programmed for (Omohundro 2008; Yudkowsky 2008). While there is reason to take this risk seriously—and indeed many do—assessing the risk cannot rely exclusively on empirical data, because no AI like this has ever existed. Characterizing AI risk thus requires expert judgment to supplement whatever empirical data is available (Baum et al. 2011). And while experts, like everyone else, are prone to make mistakes in making predictions and estimating the nature of the world, careful elicitation of expert judgment can reduce these mistakes and improve the overall risk analysis.

That said, for GCR analysis it can be especially important to remember the possibility of experts being wrong. Indeed, for very low-probability GCRs, this possibility can dominate the analysis, even when experts have wide consensus and high confidence in their conclusions, and even when the conclusions have significant theoretical and empirical basis (Ord et al. 2010). It can be similarly important to remember the possibility that experts with outlier opinions can be right (Ćirković 2012). Ordinarily, these possibilities would not merit significant attention, but the high stakes of GCR means that even remote possibilities can warrant at least some scrutiny.

A different type of analytical challenge comes from the global nature of GCRs, which makes them especially complex risks to analyze. GCRs are driven variously by the biggest geopolitical rivalries (in the case of biological or nuclear war), advanced research and development and the advantages it can confer (in the case of emerging technologies), or the entire global industrial economy (in the case of environmental collapses). Likewise, the impacts of global catastrophes depend on the resilience of global human civilization to major systemic shocks, potentially including major losses of civil infrastructure, manufacturing, agriculture, trade, and other basic institutions that enable the existence and comforts of modern civilization (Maher and Baum 2013). Assessing the extent of GCR requires accounting for the complexities of all these disparate factors plus many others.

Of course it is not possible to include every detail in any risk analysis, and certainly not for global risks. One must always focus on the most important parts. Here it is helpful to recall the high stakes associated with the most severe global catastrophes: the ones that would cause permanent harm to human civilization. While these catastrophes can be highly multifaceted, with a wide variety of impacts, the one impact that stands out as particularly important is that of permanent harm. Other impacts, and the causes of those impacts, are simply less important. A GCR analysis can focus on the permanent harm and its causes.

Climate change is an excellent example of a highly complex GCR. Climate change is caused mainly by a wide range of industrial, agricultural, deforestation, and transportation activities, which in turn are connected to a large portion of the activities that people worldwide do on a daily basis. The impacts of climate change are equally vast, affecting meteorological, ecological, and human systems worldwide, causing everything from increased storm surge to increased incidence of malaria. Each of these various impacts is important to certain people and certain ecosystems, but most of them will not make or break humanity's success as a civilization. Instead, the GCR analyst can look directly at worst-case global catastrophe scenarios, such as the possibility of temperatures exceeding the thermal limits of mammals across much of the world, in which case mammals in those regions not in air conditioning will overheat and die (Sherwood and Huber 2010). Thus a focus on GCR can make the overall analysis easier.

A subtler complexity, which GCR scholarship is only just starting to address, is the systemic nature of certain GCRs. Most GCR scholarship treats each risk as distinct: there could be a nuclear war, *or* there could be catastrophic climate change, *or* there could be an AI catastrophe, and so on. But these risks do not exist in isolation. They may be caused by the same phenomena, such as a quest for economic growth (causing both climate change, via industrial activity, and AI, via commercial technology development). Or they may cause each other, such as in the concern that climate change will lead to increased violent conflict (Gemenne et al. 2014). They may also have interacting impacts, such as if a war or other catastrophe hits a population already weakened by climate change. These various interactions suggest a systems approach to GCR analysis (Baum et al. 2013; Hanson 2008; Tonn and MacGregor 2009), just as interactions among other risks suggest a systems approach to risk analysis in general (Haimes 2012; Park et al. 2013).

Systemic effects further suggest that global catastrophe could be caused by relatively small events whose impacts cascade into a global catastrophe. Similar systemic effects can already be seen across a variety of contexts, such as the 2003 Italy power outage caused by trees hitting two power lines in Switzerland, with effects cascading across the whole system (Buldyrev et al. 2010). Just as these systems proved fragile to certain small disturbances, perhaps the global human civilization could too. Characterizing these global systemic risks can give a clearer understanding of the totality of the GCRs that human civilization now faces, and can also help identify some important opportunities to reduce or otherwise manage the risks.

Some GCR dilemmas

Unfortunately, managing GCR is not always as simple as analyzing the risks and identifying the risk management options. Some GCR management options pose deep dilemmas that are not easily resolved, even given full information about the risks. The bottom line is that, no matter how successful GCR analysis gets, society still faces some difficult decisions.

One dilemma arises from the very high stakes of GCR, or rather the very high magnitude of damages associated with permanent harm to human civilization. The high magnitude suggests that GCR reduction efforts should be prioritized over many other possible courses of action. Sometimes prioritizing GCR reduction efforts is not a significant concern, when the efforts would not be difficult or when they would be worth doing anyway, such as reducing climate change risk by improving energy efficiency. However, sometimes GCR reductions would come at a significant cost. In these cases society may think twice about whether the GCR reductions are worth it, even if the GCR reductions arguably should take priority given the high magnitude of global catastrophe damages.

An example of such a dilemma can be found for climate change and other environmental risks. Because these risks are driven by such a large portion of human activity, reducing the risks could mean curtailing quite a lot of these activities. Society may need, among other things, to build new solar and wind power supplies, redesign cities for pedestrians and mass transit, restructure subsidies for the agriculture and energy sectors, and accept a lower rate of economic growth. Individuals may need, among other things, to change their diets, modes of transportation, places of residence, and accept a simpler lifestyle. Such extensive efforts may pass a cost-benefit test (Stern 2007), especially after accounting for the possibility of global catastrophe (Weitzman 2009), but many people may still not want to do them. Indeed, despite the increasingly stark picture painted by climate change research, the issue still does not rank highly on the public agenda (Pew 2014). Should aggressive effort nonetheless be taken to reduce greenhouse gas emissions and protect the environment? This is a difficult dilemma.

A similar dilemma arises for one proposed solution to climate change: geoengineering. Geoengineering is the intentional manipulation of the global Earth system (Caldeira et al. 2013). A prominent form of geoengineering would inject aerosol particles into the stratosphere in order to block incoming sunlight, thereby reducing temperatures at the surface. While this stratospheric geoengineering could not perfectly compensate for the climatic changes from greenhouse gas emissions, it probably could help avoid some of the worst damages, such as the damages of exceeding the thermal limits of mammals.

However, stratospheric geoengineering comes with its own risks. In particular, if society stops injecting particles into the stratosphere, then temperatures rapidly rise back to where they would have been without any geoengineering (Matthews and Caldeira 2007). The rapid temperature increase could be very difficult to adapt to, potentially causing an even larger catastrophe than regular climate change. This creates a dilemma: Should society run the risk of geoengineering catastrophe, or should it instead endure the pains of regular climate change (Baum et al. 2013)? Given how bad climate change could get, this makes for another difficult dilemma.

The high stakes of GCR suggest that society should do whatever would minimize GCR, and accept whatever suffering might follow. This could mean taking aggressive action to protect the environment, or, if that does not happen, suffering through climate change instead of attempting geoengineering. Looking at the analysis on paper, it is easy to recommend minimizing GCR: the numbers simply point heavily in that direction. But in practice, this would not be an easy recommendation to make, asking many people to make such a great sacrifice. Hopefully, clever solutions can be found that will avoid these big dilemmas, but society should be prepared to make difficult choices if need be.

Another type of dilemma occurs when multiple GCRs interact with each other. Sometimes one action can reduce multiple GCRs. However, sometimes an action would reduce one GCR while increasing another. This poses a dilemma between the two GCRs, a risk–risk tradeoff (Graham and Wiener 1995). A good example of this type of dilemma is nuclear power. Nuclear power can help reduce climate change risk by shifting electricity production away from coal. However, nuclear power can also increase nuclear war risk by enabling nuclear weapons proliferation. This dilemma is seen most clearly in ongoing debates about the nuclear program of Iran. While Iran claims that its program is strictly for peaceful electricity and medical purposes, many observers across the international community believe that Iran is using its program to build nuclear weapons.

Given the risks from climate change and nuclear war, should nuclear power be promoted? How much should it be promoted, and in what circumstances? Resolving these dilemmas requires quantifying and comparing the two GCRs to identify when nuclear power would result in a net decrease in GCR. Unfortunately, at this time GCR research has not quantified climate change and nuclear war risk well enough to be able to make the comparison and reach conclusions about nuclear power. Meanwhile, in circumstances in which nuclear power would not create a nuclear weapons proliferation risk, such as for countries that already have nuclear weapons or clearly do not want them, nuclear power probably would bring a net GCR reduction. This conclusion brings up the first type of dilemma—the general sacrifice for GCR reduction—where nuclear power raises concerns about industrial accidents like Chernobyl or Fukushima. Such accidents are "only" local (or regional) catastrophes, but they are nonetheless plenty large enough to weigh heavily in decision making.

Conclusions: a research agenda

Regardless of whether GCR reduction should be prioritized above everything else, it is clear that GCR is an important issue and that reducing GCR merits some effort. The big research question then is, which efforts? That is, what are the best, most effective, most desirable ways to reduce GCR? Unfortunately, the GCR research community has not yet made significant advances to answering this vital question. A new research agenda is needed for it.

We believe that GCR research is most helpful at guiding GCR reduction efforts when the research covers all the major risks and risk reduction options in a consistent, transparent, and rigorous manner. It should to include all the major risks and risk reduction options in order to identify which ones are most important and most worth pursuing. Analyzing each risk in isolation fails to account for their various systemic interactions and prevents evaluating risk–risk dilemmas like that posed by nuclear power. In contrast, integrating all the risks and risk reduction options into one assessment can help decision makers identify the best options. Similar integrated assessments have been done for a variety of other topics, such as the popular economic assessments of climate change (e.g. Nordhaus 2008). Something along these lines, adapted for the particulars of GCR, would be of great value to GCR reduction decision making.

An integrated assessment of GCR poses its own analytical challenges. The particulars of different GCRs can be quite different from each other. Fitting together e.g. a climate model, an epidemiological model, and a technological development model would not be easy, nor would filling in the important gaps that inevitably appear between the models. Each GCR is full of rich complexity; the full system of GCRs is more complex still. This makes it even more important to focus on the most important aspects of GCR, lest the analysis get bogged down in details.

It is equally important for the analysis to focus on risk reduction options that are consistent with what attentive decision makers are in a position to do. Otherwise the analysis risks irrelevance. While this is true for any analysis, it is especially true for GCR. The global scale of GCR makes it tempting for analysis to ignore details that seem small relative to the risk but are important to decision makers, and also to entertain risk reduction options that perform well in a theoretically ideal world, "if only everyone would do that." Furthermore, the high stakes of GCR makes it tempting for analysis to recommend rather drastic actions that go beyond what most people are willing to do. It is thus that much more important for GCR analysis to remain in close touch with actual decision makers, to ensure that the analysis can help inform actual decisions.

Despite these challenges, we believe that such a research agenda is feasible and can make an important contribution to society's overall efforts to reduce GCR. Indeed, the future of civilization could depend on it.

References

Adams, F.C. and Laughlin G., 1997. A dying universe: the long-term fate and evolution of astrophysical objects. *Reviews of Modern Physics*, 69(2), 337–72.

Atkinson, A., 1999. *Impact Earth: Asteroids, Comets and Meteors: The Growing Threat*. London: Virgin.

Barrett, A.M., Baum, S.D. and Hostetler, K.R., 2013. Analyzing and reducing the risks of inadvertent nuclear war between the United States and Russia. *Science and Global Security*, 21(2), 106–33.

Baum, S.D., Goertzel, B., and Goertzel, T.G., 2011. How long until human-level AI? results from an expert assessment. *Technological Forecasting & Social Change*, 78(1), 185–95.

Baum, S.D., Maher, T.M. Jr., and Haqq-Misra, J., 2013. Double catastrophe: intermittent stratospheric geoengineering induced by societal collapse. *Environment, Systems, and Decisions*, 33(1), 168–80.

Bailey, R.T., 1997. Estimation from zero-failure data. *Risk Analysis*, 17(3), 375–80.

Beckstead, N., 2013. *On the Overwhelming Importance of Shaping the Far Future*. Doctoral Dissertation, Department of Philosophy, Rutgers University.

Bossert, W., Sprumont, Y., Suzumura, K., 2007. Ordering infinite utility streams. *Journal of Economic Theory*, 135, 579–89.

Bostrom, N., 2002a. Existential risks: analyzing human extinction scenarios and related hazards. *Journal of Evolution and Technology*, 9(1), n.p.

Bostrom, N., 2002b. *Anthropic Bias*. New York: Routledge.

Bostrom, N. and Ćirković, M., 2008. *Global Catastrophic Risks*. Oxford, UK: Oxford University Press.

Buldyrev, S.V., Parshani, R., Paul, G., Stanley, H.E., and Havlin, S., 2010. Catastrophic cascade of failures in interdependent networks. *Nature*, 464, 1025–28.

Caldeira, K., Bala, G., and Cao, L., 2013. The science of geoengineering. *Annual Review of Earth and Planetary Sciences*, 41, 231–56.

Ćirković, M.M., 2002. Cosmological forecast and its practical significance. *Journal of Evolution and Technology*, 7, n.p.

Ćirković, M.M., 2012. Small theories and large risks—Is risk analysis relevant for epistemology? *Risk Analysis*, 32(11), 1994–2004.

Ćirković, M.M., Sandberg, A., and Bostrom, N., 2010. Anthropic shadow: observation selection effects and human extinction risks. *Risk Analysis*, 30(10), 1495–506.

Crevier, D., 1993. *AI: The Tumultuous History of the Search for Artificial Intelligence*. New York: Basic Books.

Gemenne, F., Barnett, J., Adger, W.N., Dabelko, G.D., 2014. Climate and security: evidence, emerging risks, and a new agenda. *Climatic Change*, 123(1), 1–9.

Graham, J.D., and Wiener, J.B., eds, 1995. *Risk vs. Risk: Tradeoffs in Protecting Health and the Environment*. Cambridge, MA: Harvard University Press.

Haimes, Y.Y., 2012. Systems-based guiding principles for risk modeling, planning, assessment, management, and communication. *Risk Analysis*, 32(9), 1451–67.

Hanley, J. A. and A. Lippman-Hand (1983). If nothing goes wrong, is everything all right? *Journal of the American Medical Association*, 249(13), 1743–45.

Hanson R., 2008. Catastrophe, social collapse, and human extinction. In N. Bostrom, M.M. Ćirković, eds, *Global Catastrophic Risks*. Oxford, UK: Oxford University Press. pp.363–78.

Hempsell, C.M., 2004. The potential for space intervention in global catastrophes. *Journal of the British Interplanetary Society*, 57, 14–21.

Kaku, M., 2005. Escape from the universe. *Prospect Magazine*, February 20. www.prospectmagazine.co.uk/magazine/escapefromtheuniverse

Maher, T.M. Jr., and Baum, S.D., 2013. Adaptation to and recovery from global catastrophe. *Sustainability*, 5(4), 1461–79.

Matheny, J.G., 2007. Reducing the risk of human extinction. *Risk Analysis*, 27(5), 1335–44.

Matthews, H.D., and Caldeira, K., 2007. Transient climate–carbon simulations of planetary geoengineering. *Proceedings of the National Academy of Sciences*, 104, 9949–54.

McCorduck, P., 2004. *Machines Who Think: 25th Anniversary Edition*. Natick, MA: A.K. Peters.

Nagel, T., 1986. *The View from Nowhere*. Oxford, UK: Oxford University Press.

Ng, Y.-K., 1991. Should we be very cautious or extremely cautious on measures that may involve our destruction? *Social Choice and Welfare*, 8, 79–88.

Nordhaus, W.D., 2008. *A Question of Balance: Weighing the Options on Global Warming Policies*. New Haven, CT: Yale University Press.

Omohundro, S.M., 2008. The basic AI drives. In P. Wang, B. Goertzel and S. Franklin, eds, *Artificial General Intelligence 2008: Proceedings of the First AGI Conference*. Amsterdam: IOS Press. pp.483–92.

Ord, T., Hillerbrand, R., and Sandberg, A., 2010. Probing the improbable: methodological challenges for risks with low probabilities and high stakes. *Journal of Risk Research*, 13(2), 191–205.

Park J., Seager, T.P., Rao, P.S.C., Convertino, M., and Linkov, I., 2013. Integrating risk and resilience approaches to catastrophe management in engineering systems. *Risk Analysis*, 33(3), 356–67.

Pew, 2014. *Thirteen years of the public's top priorities*. Washington, DC: The Pew Research Center, 27 January. http://www.people-press.org/interactives/top-priorities/

Posner, R., 2004. *Catastrophe: Risk and Response*. Oxford, UK: Oxford University Press.

Quigley, J. and Revie M., 2011. Estimating the probability of rare events: addressing zero failure data. *Risk Analysis*, 31(7), 1120–32.

Rawls, J., 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Rees, M., 2003. *Our Final Century: Will the Human Race Survive the Twenty-First Century?* Oxford, UK: William Heinemann.

Robinson, L.A., 2007. How US government agencies value mortality risk reductions. *Review of Environmental Economics and Policy*, 1(2), 283–99.

Sagan, C., 1983. Nuclear war and climatic catastrophe: some policy implications. *Foreign Affairs*, 62(2), 257–92.

Sagan, S.D., and Waltz, K.N., 2013. *The Spread of Nuclear Weapons: An Enduring Debate,* third edn. New York: W.W. Norton.

Sherwood, S.C., and Huber, M., 2010. An adaptability limit to climate change due to heat stress. *Proceedings of the National Academy of Sciences*, 107, 9552–5.

Stern, N.H., 2007. *The Economics of Climate Change: The Stern Review*. Cambridge, UK: Cambridge University Press.

Tonn, B., MacGregor, D., 2009. A singular chain of events. *Futures*, 41, 706–14.

Weitzman, M.L., 2009. Structural uncertainty and the value of statistical life in the economics of catastrophic climate change. *Review of Economics and Statistics*, 91(1), 1–19.

Yudkowsky, E., 2008. Artificial intelligence as a positive and negative factor in global risk. Global Catastrophic Risks. In N. Bostrom, M.M. Ćirković, eds, *Global Catastrophic Risks*. Oxford, UK: Oxford University Press. pp.308–45.

Index

Page numbers in **bold** refer to figures, page numbers in *italic* refer to tables.