

Ivan Jureta

# Analysis and Design of Advice

 Springer

# Analysis and Design of Advice



Ivan Jureta

# Analysis and Design of Advice

Ivan Jureta  
University of Namur (FUNDP)  
Louvain School of Management  
Rempart de la Vierge 8  
5000 Namur  
Belgium  
ivan@jureta.net

ISBN 978-3-642-18058-3 e-ISBN 978-3-642-18059-0  
DOI 10.1007/978-3-642-18059-0  
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011923527

ACM Computing Classification (1998): D.2, H.4, K.4, J.1

© Springer-Verlag Berlin Heidelberg 2011

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

*Cover design:* deblik, Berlin

Printed on acid-free paper

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

*For my parents, Ljiljana & Jozo*



# Preface

To the advisor, an analysis of advice should hint at what to suggest and how to do so. To those who receive advice, it should help them decide how to respond to suggestions.

How we analyze advice affects fundamentally the way we understand the state of affairs as it is and might be, who decides these questions, and how we choose what can, ought to be, and is done. It influences what we believe and do, our perception of whom is of similar or different knowledge, authority, and our conception of desirable or undesirable positions within culture, politics, economics, and industry.

This book offers a general way to analyze advice. The analysis applies regardless of what advice is about and whom it comes from or needs to be given to, and is in this sense general. The analysis is made of two intertwined parts, a conceptual analysis and the analysis of the rationale of advice.

Communicated in an attempt to more or less precisely refer to concepts, objects, and the relations between these, advice aims to convey information about conditions, events, and situations to influence decisions. *Conceptual analysis* concentrates on the difficulties in understanding advice and making it understandable in intended ways, why these issues arise, how they manifest themselves, and what to pay attention to when choosing what to advise and how to do so. The rationale of advice is the assumptions on which it is based and the goals which it should achieve for the advisor or the decision-maker who receives it. *Analysis of the rationale of advice* aims to define and describe assumptions and goals, and from there determine whether advice is appropriate within a given decision situation, how robust it is with regards to criticism that may be directed toward it, and how it relates to the choice expected of the decision-maker who receives it.

The introduction is followed by a chapter that develops and defends the argument that the way an individual analyses advice influences the way he coordinates with others, and thereby his position within mechanisms of coordination, such as markets in economics and polyarchies in politics. Chapters 3 and 4 are dedicated to the conceptual analysis of advice and Chaps. 5 and 6 to the analysis of the rationale of advice.

The premise that giving advice is a *design problem* is central to this book. This means that advice is seen as an artifact, as information communicated by an advisor to a decision-maker. This information can and ought to be designed, its properties



decided on the basis of an understanding of the context in which it is given, the decision problem and the decision-maker it targets, what the decision-maker may know or ignore, and meet the objectives of the decision-maker and of the advisor. The design of advice is a goal-oriented activity that can be approached in a more systematic way once there is a grasp of the unavoidable difficulties commonly encountered both by the advisor and by the decision-maker, as well as of the feasible responses thereto. For the advisor, the analysis of advice is a method for the resolution of the problem that he faces, which is what advice to give and how to do it. For the decision-maker, it is a method to evaluate the advice that he receives.

This book was originally intended for undergraduate and graduate students of management science, as many of them go on to become professional advisors. Fields such as operations research and finance already teach them the methods that they can apply to make recommendations, especially when they can collect or already have quantitative data. These decision-making situations are, so to speak, well structured: they are the ones in which much rigorous thought was already invested, for which agreed-upon instruments of measurement have been defined and used to desirable results, where a solid body of knowledge is already available and ought to be applied to come up with a recommendation. However, many decision situations are not as convenient. The intention in this book is to offer a general method of analysis that applies to the giving and receiving of advice when the decision problems are not well structured and in which there are variously imprecise, unclear, incomplete, or conflicting qualitative information.

The reader is not expected to have the specific knowledge that management science students acquire before they are confronted with the issues raised here. By aiming for a rigorous discussion, commitment is expected for the ideas to be understood to the point at which they can be used, criticized, and hopefully built upon and revised. This is especially the case toward the end of the book, when mathematical models of advice are introduced. They are, as most of the ideas here, built from grounds up, in small and simple steps.

October 2010

*Ivan Jureta*

# Acknowledgments

Marta Coelho's and Bernd Irlenbusch's lectures at the London School of Economics introduced me to the research on decision theory and decision analysis. I became interested in requirements engineering through three very different lectures on the engineering of information systems, one Ian Angell's, again in London, the other two at Université catholique de Louvain, one with Philippe Wilmès, the other with Manuel Kolp. My ideas on the analysis of advice took their preliminary shape at that time. They are inspired by, and define themselves in relation to the contributions in these fields.

Four colleagues merit particular mention. It is mainly with them that I have discussed and published many of the ideas that have preceded and underlie this book. Pierre-Yves Schobbens helped me discover the many lines of inquiry in requirements engineering and, in particular, their relationships to artificial intelligence. Stéphane Faulkner was my mentor in research, and his enthusiasm and support seem to have no limits. I met John Mylopoulos at the University of Trento in 2007. We then initiated a conversation about the foundations of the requirements engineering field which is ongoing still. His advice was always and still is exceptional. Alex Borgida introduced me to mathematical logics from the standpoint of those who make them.

Many others have helped. I wish to thank Marco Saelens, Esteban Zimányi, Neil Ernst, Sotirios Liaskos, and Patrick Heymans. Ralf Gerstner at Springer offered many relevant suggestions that have helped me to improve this book. Others will, I hope, forgive me for having omitted them. Some of them are the undergraduate and graduate students at the management department of the University of Namur and Université catholique de Louvain, who attended the lecture "Decision Making & Requirements Engineering" in 2009 and on whom I tried out the ideas developed in this book.

The analysis of advice presented here arises from and resembles in its assumptions, concepts, and models my research in requirements engineering. Discussions presented in this book developed in parallel with that research. I did most of my work in requirements engineering at the University of Namur, at the departments of management and computer science. Parts were elaborated during stays at University of Trento, University of Toronto, and Carnegie Mellon University. Most, if not all of it was presented and published between 2005 and 2010 at

various international conferences and journals for requirements engineering, conceptual modeling, and information systems engineering. Many anonymous reviewers and audiences have offered relevant remarks and suggestions. They have helped shape those publications and the ideas in this book. I should single out those at IEEE *International Requirements Engineering Conference*, *International Conference on Conceptual Modeling*, *Requirements Engineering Journal*, and *Applied Ontology* journal. Collège Interuniversitaire pour les Sciences du Management initially, then Fonds de la Recherche Scientifique–FNRS funded my research, giving me the freedom to pursue these ideas.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Coordination and Advice</b>	<b>7</b>
2.1	When There Is None	9
2.2	From Division of Labor to Dependence on Advice	11
2.3	Autonomy and Coordination	15
2.4	Coordination Through Advice	18
2.5	Advice in Political Coordination	19
2.5.1	Dictatorship and Consent	20
2.5.2	Under Democracy	21
2.5.3	Advice and Meta-Advice in Politics	25
2.6	Advice in Economic Coordination	34
2.6.1	Central Planning	35
2.6.2	Market	40
2.7	Homo Follis	43
<b>3</b>	<b>Conceptual Analysis of Advice</b>	<b>47</b>
3.1	Rigorous Definition	51
3.1.1	Ostensive and Intensional Definitions	51
3.1.2	Primitive Terms	54
3.1.3	Criteria from Standard Theory	58
3.2	Signs to Objects and Back, via Concepts	61
3.2.1	Signs	62
3.2.2	Objects and Concepts	63
3.2.3	Relating Signs, Objects, and Concepts	65
3.2.4	Sings, Objects, and Concepts, in Advice	78
3.3	Ontology and Primitive Terms	82
3.3.1	Ontology in Philosophy	83
3.3.2	On the Engineering of Ontologies	89
3.4	Advice, Defined	97
3.4.1	Initial Ontological Commitments	97
3.4.2	Advice in Communication, Communication as Action	99

3.4.3	Essential Properties of and Identity Criteria for Advice .....	104
3.4.4	What Advice Is Not, but May Refer To.....	107
3.5	Relativist's Conceptual Analysis .....	110
<b>4</b>	<b>Interpretation of Advice.....</b>	<b>115</b>
4.1	Open Reference .....	118
4.2	Vague Advice.....	123
4.3	Referent Lost .....	132
<b>5</b>	<b>Kinds of Advice .....</b>	<b>141</b>
5.1	Classification from a Model of Choice.....	143
5.2	Perfect and Bounded Rational Choice .....	146
5.2.1	Ontology of Decision Analysis .....	150
5.2.2	Ontology of Choice in Organized Anarchies .....	152
5.2.3	Ontology of Intervowen Organisational Choice.....	154
5.3	Intolerance for Substitutes .....	157
5.3.1	Probability Intolerance .....	160
5.3.2	Utility Intolerance .....	168
5.4	Decision Information .....	172
5.4.1	Revealed Intentional States.....	173
5.4.2	Decision Information Ontology .....	183
5.4.3	Synthesis of the Decision Information Ontology.....	194
5.5	Taxonomy of Advice.....	196
5.5.1	Whose Explanations?.....	196
5.5.2	Specialization of the Concept of Advice .....	198
5.6	Reinterpreting Advice .....	202
<b>6</b>	<b>Advisor's Problem and Its Solutions.....</b>	<b>207</b>
6.1	Advice from Simple Explanations and Predictions.....	210
6.2	Models of Advice: An Overview .....	213
6.2.1	Introductory Example .....	213
6.2.2	Overview of the Framework for the Modeling of Advice .....	219
6.3	Modeling Language.....	224
6.3.1	AML2 .....	225
6.3.2	AML1 .....	236
6.3.3	AML.....	245
6.3.4	A-nets .....	254
6.3.5	Why the Interest in Structure? .....	257
6.4	Formulation of the Advisor's Problem .....	259
6.5	Solving the Advisor's Problem .....	265
6.6	Criteria for a Robust Solution .....	273

6.7	Criteria for a Clear Solution .....	280
6.8	Criteria from Empirical Evidence .....	281
<b>7</b>	<b>Perspectives</b> .....	<b>285</b>
	<b>References</b> .....	<b>287</b>
	<b>Index</b> .....	<b>297</b>



# Chapter 1

## Introduction

Advice recommends what to think; through thought, what to choose; and via choices, how to act. Advice is information that moves by communication, from advisors to the recipient of advice, individuals whose choices need to be influenced. It is a tool for coordination, for its acceptance means predictability, necessary toward the achievement of goals that remain beyond the reach of an individual. Advice plays a central role in all affairs requiring coordination.

Analysis and design of advice should be distinguished, as they have different purposes.

The aim of *analyzing advice* is, roughly speaking, to identify the thoughts, choices, and actions that advice targets, and recommends or counters, as well as to hypothesize the consequences that its acceptance or rejection may have. The result of such an analysis is to choose whether to accept advice or reject it.

To *design advice*, it is necessary to see it as an artifact, a product that can be carefully designed, in some sense as bridges or aircrafts are: design of advice aims to choose the purpose of advice, then determine the properties that its content, form, and method of delivery should have for advice to realize that purpose.

What this book aims to accomplish is to define a *general method for the analysis and design of advice*.

A method for analysis and design is *general* if it applies across topics and concerns; otherwise, the method is specific: it focuses on a particular domain, such as economics, finance, and politics, or goes deeper within the subdivisions of each field, e.g., development economics, corporate finance, foreign affairs. Such a focus is not the aim here. Rather, the book concentrates on the production and consumption of advice independently of a particular field.

The construction of a *general* method for the analysis and design of advice is necessarily a *theoretical* exercise in *method engineering*. This emphasis on theory contrasts to the practice of producing and consuming advice, both of which are exceptionally accessible. It is effectively enough to be competent in some form of communication to be able to give and take advice. This accessibility has nevertheless nothing to do with either the *procedural quality* or the *outcome quality* of advice. That something can be done guarantees neither that it is (always) done well nor that it will (always) give the desired results.



*Procedural* quality concerns how advice was designed *before* it is given or how it is analyzed *before* it is accepted or rejected. *Outcome* quality focuses on how appropriate – roughly, good or bad – advice proved to the individual who received it. Judgment of outcome quality is highly subjective, which makes it easier to influence how advice is designed before delivery, than to affect how it will be evaluated after reception. Same problems apply to the design of any product: engineering, manufacturing, delivery, and advertising can be controlled better than the emotional reaction of a particular consumer.

The general method of analysis and design focuses on improving the procedural quality of advice, not directly its outcome quality. Sometimes the two may be related, but the precise conditions under which this would be the case remain elusive. Just as companies hope that care in the engineering, manufacturing, delivery, and advertising would make for a product that gives a favorable impression to its users, so does the designer of advice expect the careful design of advice to elicit the desired reaction, that is, the thoughts, choices, and actions that advice was made to recommend.

To view advice as an artifact, as a product that must be designed, is to see the production and analysis of advice as a *design problem*. To offer a general method for the analysis and design of advice amounts thus to give a method for the *resolution of that design problem*.

The emphasis on theory in this book may surprise, so a myth must be dispelled right away. There is a tendency to think that practice and experience must precede theory when methods are constructed to solve design problems. This often mistakenly translates into a dislike for theoretical discussions. The usual arguments are that theoretical considerations prove irrelevant to practice, that they are remote, and that what ultimately matters is only what is done: the choices made, the ensuing actions, and the observed effects. It is also usually said that “talk is cheap,” and to the extent that theory does take the form of talk (one which nevertheless leaves a paper or otherwise trail), theory remains disconnected from practice. It is then common to hear that a method was built out of intuition adjusted through trial and error, and there is, really, no need for much theorizing. This kind of method engineering can be good enough, but there is no need to be content with such an approach when a better one is feasible. Theoretical studies that precede and follow practice, record and study unavoidable problems and relevant responses, both of which are often independent of a field’s specifics. A myopic emphasis on doing, rather than at least systematically thinking before doing, is thus at best lazy and careless.

This book constructs the general method for the analysis and design of advice through theoretical discussions informed as much as possible by empirical research. This is a rather unsurprising approach. What is unusual is the theoretical backdrop against which the discussions are set: ontology engineering, conceptual analysis, and (some rather simple) artificial intelligence. Ontology engineering and conceptual analysis offer theoretical tools through which advice is defined and distinguished from other kinds of information. The result is a definition of the very concept, advice, the instances of which have to be analyzed and designed. As advice itself is information communicated under particular conditions, these tools serve

in the analysis and design of the information that advice should convey. Artificial intelligence and, in particular, its branches focused on knowledge representation and reasoning give the theoretical tools to represent, to model advice in forms that both facilitate some particular kinds of analysis and design tasks, and can be amenable to automated analysis, that is, the use of computers to evaluate advice. While Chap. 2 of the book emphasizes the role of advice in the coordination of people, Chaps. 3, 4 and 5 describe and apply the tools of ontology engineering and conceptual analysis to analyze and design advice. Some basic theoretical tools of artificial intelligence are constructed in Chap. 6, namely a mathematical formalism is defined that can represent advice and allows analyses that are beyond the scope of both ontology engineering and conceptual analysis.

To the extent that advice is given to influence decisions, it would have been expected that classical decision theory, one developed mainly in economics, plays a key role here. This is not the case primarily for two reasons. One is the rejection, in this book, of assumptions of perfect rationality. As it is argued in Chap. 4, such assumptions would make the analysis and design of advice entirely uninteresting. The assumptions are not realistic in relation to the production and consumption of advice, and they are not accepted. Second, and more importantly, classical decision theory relies to a considerable extent on numerical, quantitative estimates of probability and utility. Actual production and consumption of advice happens in situations in which *relevant* numerical estimates are either impossible to obtain or are very hard to come by. This entails an intolerance in this book to the *quantitative* treatment of uncertainty and desirability, in favor of a qualitative one: mathematical logic is thus the tool of choice, rather than probability and utility functions. This is clear in Chap. 6, where the aim is to define as precisely as feasible the design problem that advisor is expected to solve, that is, the problem of deciding the properties of the content, form, and delivery of advice.

The departure from classical decision theory here goes together with a departure from its revisions that accommodate violations of perfect rationality assumptions, but keep a quantitative treatment of uncertainty and desirability. Such a stance also puts an evident distance between this book and the research on decision analysis, and this for the very same reasons that classical decision theory is set aside: emphasis on quantitative estimates of utility and probability simply requires unrealistic assumptions about the production and consumption of advice.

Casting aside well-known conceptual tools and assumptions, the change of theoretical backdrop led to the development and justification of new concepts and methods in this book. They are presented and articulated as the following fundamental components of the general method for the analysis and design of advice:

1. *The ontology of choice and advice.* The ontology defines what advice is, how it relates to decisions, and what basic kinds of advice there are. The ontology identifies the kinds of information that are used and produced in the analysis and design of advice.
2. *The conceptual analysis of advice.* Given the information identified by the ontology of choice and advice, the conceptual analysis concentrates on the intended

use of this information, and its relationships to potential choices within the decision situation in which advice is given. Conceptual analysis aims to give advice the content, form, and method of delivery that would make it understandable in intended ways.

3. *The analysis of the rationale of advice.* The rationale of advice are the reasons that justify it having a particular content, form, and method of delivery. This analysis involves the finding of the assumptions on which advice relies, the goals which it should satisfy for the advisor and the decision-maker who receives it, and the relationships between the assumptions and the goals. The aim is to make explicit and clarify these assumptions and goals, and from there decide whether the content, form, and method of delivery of advice are appropriate within a given decision situation, how robust advice is with regards to criticism that may be directed toward it, and how it relates to the choice expected of the decision-maker who receives it.

The three components above form the general method for the analysis and design of advice. Ontology identifies and categorizes information that is input to the conceptual and rationale analyses, and categorizes the information produced by these analyses. Conceptual and rationale analyses transform that information to determine which advice to give and how to do so.

Through the development of the general method, this book pursues six aims:

- To advance a conception of advice as an artifact that can be designed.
- To present, illustrate, define, and justify the components of the general method for the analysis and design of advice.
- To present and argue for an approach to the analysis of advice and its interaction with decisions that is alternative to classical decision theory and decision analysis.
- To argue for and present the advantages and problems of developing a *qualitative* (symbolic, that uses mathematical logic) analysis of advice and choice, that does not give a central role to numerical estimates of probability and utility, but favors symbolic representation of qualitative information, and the automation of the analysis of that information through already available computational means developed mainly for knowledge representation and reasoning, within the artificial intelligence field.
- To contribute to the theoretical understanding of the production and consumption of advice, and the interaction between advice and choice.
- To argue for the use of contributions in ontology engineering, conceptual modeling and analysis, and knowledge representation and reasoning in artificial intelligence for the systematic analysis of decision situations, and thereby advocate the use and further development of these contributions for the analysis of phenomena that clearly fall within management science, in this case the production and consumption of advice.

These six aims are deeply connected. As soon as some basic properties of the production and consumption of advice are recognized, quantitative estimates of

probability and utility cannot be used to handle uncertainty and desirability in the analysis and design of advice. It consequently becomes necessary to depart from classical decision theory and decision analysis, develop new concepts to capture information about advice and choice, establish new relations between the concepts, and define new mathematical formalisms for their modeling and analysis. Together, the new concepts, relations, and formalisms form a general, qualitative method for the analysis and design of advice and choice.



## Chapter 2

# Coordination and Advice

Three processes shape developed contemporary societies and give the analysis of advice its central role: the advancing division of labor, the widening field of personal decision, and the decreasing cost of distributing information.

Division of labor along well-delimited tasks in production, the specialization of work came to prominence in the time before, and progressed considerably throughout the eighteenth and nineteenth century Industrial Revolution. With it came the specialization of knowledge, reflected in the fragmentation of scientific inquiry, the organization of educational systems along specialties, and societies that increasingly value expertise, the deep, but narrow knowledge. To make decisions that fall within the limits of his or her own specialty, the citizen draws on the expert knowledge that he or she acquires. As specialties continually narrow and deepen, it becomes increasingly difficult to admirably navigate and command more than a single area of expertise. The consequence is that, when facing a decision outside his area of competence, the specialist will have to rely on advice. The narrower and deeper the specialization, the greater is the number of situations in which it is advice that informs or misinforms one's decisions, and the more critical it is, therefore, to rigorously analyze it.

The field of personal decision widens through pursuits of laudable freedoms of thought, speech, and action and the increasing reliance on reason and science for insight. Although still beset by superstitions, taboos, and myths, contemporary societies nevertheless are incomparably less so than their tribal, or even recent predecessors. Instead of following the rules set by chiefs, kings, and mystics, the present-day citizen is expected to approach very many choices autonomously and responsibly. Nowhere is this easier to observe than in democracies, where calls and procedures to decide who governs are inscribed in the foundations of law. The wider the field of personal choice, the more and the varied are the decisions that the citizen is asked to make. Constrained by time, attention, and the regularities of economics, the specialist citizen will seek advice and be confronted to the difficult question of how to respond to it, of whether to accept it.

Scientific and engineering advances of the twentieth century, and the ensuing new technologies continue to change ways in which information is produced and distributed. The cost to store and to copy text, sound, images, video are almost an insignificant fraction of what they were only a decade or more ago. Means to access

information have correspondingly decreased in cost, while those to push information to individuals have multiplied. Advice has never before been as accessible, varied, and unavoidable. Its quality does not arise from availability, making rigorous analysis crucial both to produce it for others and to consume it.

Rigorous discussion of a common phenomenon typically starts with the observation that the topic of interest is prevalent in contemporary society. There is really no need here to curb ambitions to current affairs only. As soon as there is ability to communicate and need to coordinate, there is giving and taking of advice. From the simplest cases of telling a child what to do, up to recommending courses of action to heads of state, determining what advice to give or take was, is, and will continue to be a pressing concern.

Advice-giving is a profession to some and has been so at least for as long as there were counselors to tribal chiefs and kings. Early professional advisors are depicted in some of the most prominent artworks. The king of Denmark in Shakespeare's *Hamlet* is advised by Polonius, whom William Hazlitt, an early eighteenth century English writer and literary critic, calls officious, garrulous, and impertinent, or in other words, intrusive, full of trivial conversation, and irrelevant. Not all of his advice is such, as when he tells his son Laertes in the first act of the third scene, "neither a borrower nor lender be; for loan oft loses both itself and friend; and borrowing dulls the edge of husbandry." Advice may be a gift: instead of presenting his prince, Lorenzo de' Medici the Magnificent with arms, cloth of gold, or precious stones, Niccolò Machiavelli offers him recommendations on governance in *The Prince*. It is a twisted gift, for "good counsel, from whoever it comes, must result from the prudence of the prince, and not his prudence from the good counsel" as "each counselor will consider his own interests, and [the] prince will not be able to correct them or even recognize them" (Machiavelli 2009, Sect. 22).

It is not in tribes and monarchies, but in the twentieth century republics that the institutionalization and professionalization of advice-giving took place. Advisors remained, while chiefs, kings, and princes were replaced by government officials, captains of industry, and robber barons.

Recognition of a relevant topic then usually leads to a look at the common readings of the word, its loose, but intuitively accessible definition that tends to require no particular background knowledge from the reader. The *Oxford English Dictionary* gives seven such readings. The earliest goes back to the thirteenth century, saying advice is the way in which a matter is looked at or regarded, an opinion or judgment. The word was later used to designate forethought, prudence, or wisdom; the weighing of opinions, deliberation, or consultation; opinion given or offered as to action; the result of consultation; a decision of a deliberative body; or, rather generally any information given. In some cases, advice referred to a provision for, endowment, advancement, which comes from a French use of *avis*, traced back to the fifteenth century; yet in others, it is a prefix, as in "advice-boat," which is a seventeenth and eighteenth century word for a vessel used to scout the sea and bring information back. *WordNet*, a lexical database of English, instead gives one reading of the word, namely that advice is a *proposal for an appropriate course of action*.

This is the informal reading that will serve throughout this chapter, and that certainly agrees with the perspectives given in traditional dictionaries.

A dictionary definition does little more than feebly scratch the surface of the concept of advice, and does nothing to inform any analysis of actually given or received advice. One purpose of this text as a whole is to discuss what advice is, and distinguish it from what it is not. Based on such distinctions, the other purpose is to construct a general way to analyze advice, and to see what kinds of insight such analysis can provide. While the third chapter will discuss a more elaborate definition of advice, this second chapter will use the commonsense understanding of advice as a recommendation on a course of action. The purpose in doing so is to avoid at the outset a technical definition when substantiating some important claims; the technical definition offered later remains, of course, consistent with the commonsense readings.

The key claim of this chapter is that how an individual analyses advice influences his position to and within the mechanisms of coordination in the realms of politics, economics, law, to mention a few. To support this claim, the following line of argument is taken. It is first observed that the only form of society in which the individuals rely on no advice is a society of one (Sect. 2.1). Once there is division of physical and intellectual labor, there is specialization. This in turn requires that the work of the specialists be coordinated, whereby coordination will require that advice is dispensed (Sect. 2.2). Now, we can only argue that advice will be a tool of coordination if it does play a role in the decision-making process of the individual, which leads us to consider briefly the interaction between advice and choice, as it has been observed in laboratory settings, in experiments (Sect. 2.3). To see advice as a tool of coordination leads us to discuss the role of advice in coordination mechanisms. A considerable part of this chapter is dedicated to the discussion of the use of advice for coordination within politics and economics (Sects. 2.4–2.6). This finally leads us to close the chapter by reiterating the importance for the individual of the independent assessment of advice for his relationship to coordination mechanisms (Sect. 2.7).

## 2.1 When There Is None

With warm, dry summers and cool, wet winters, no deadly predators to speak of, freshwater sources, edible berries and accessible and plentiful fishing spots, the least attractive trait of the island of Aguas Buenas may well be its distance from the nearest large landmass. The island is one of three that form the Juan Fernández archipelago, situated some 700 km west off the Chilean coast in the South Pacific, at about the same latitude as Santiago, Buenos Aires, and Montevideo. Its 15 min of fame came in late 2005, when the popular press reported an archaeological find. Excavations revealed that the island hosted an European occupant, a Scottish seaman, who spent more than 4 years there in the early eighteenth century.

It is apparently by his own choice that Alexander Selkirk was marooned in 1704 on Aguas Buenas, which was uninhabited at the time. Doubting the seaworthiness of



his ship, he asked the captain for permission to remain on the island. He did survive the ordeal, and is considered as the real-life castaway who inspired Daniel Defoe's novel on the adventures of Robinson Crusoe. Chance favored him more than once in his misfortune. Not only did he choose a rather convenient island to be marooned on, but was proved right about his ship. It sank shortly after Selkirk was left on the island, killing many of his shipmates.

It is ironic that a castaway has so much autonomy in choice, yet so little to choose from. Which course of action will be taken is entirely independent of fellow man, for none is in sight. This seems not without its benefits. As William Cowper, an eighteenth century English poet, observed in *The Solitude Of Alexander Selkirk*, the castaway rules his domain unquestioned:

I am monarch of all I survey,  
My right there is none to dispute;  
From the center all round to the sea  
I am lord of the fowl and the brute.

Although of seemingly unconstrained choice, a castaway is forced to enact an autocratic monarchy. A dictator, it is his remote and solitary condition that makes him a self-appointed ruler and forces him to decide alone on the use of resources at his disposal. His is, nevertheless, an unfortunate domain to rule.

If Selkirk were educated in the eighteenth century Scotland, he would have benefited from one of the most advanced systems of education in Europe at the time, with tax-subsidized and state-regulated schooling. Once he had set foot on the island and had explored the surroundings to ascertain the absence of immediate danger, the educated castaway would undoubtedly think of Plato's *Republic*. "Society originates," he would recall, "...because the individual is not self-sufficient, but has many needs which he can't supply himself...Quantity and quality are therefore more easily produced when a man specializes appropriately on a single job for which he is naturally fitted, and neglects all others" (Plato 2003, pp. 55–56.) If Selkirk were further interested in new economic ideas of his times, this line of thinking would remind him of William Petty. After studying Dutch ship building in the seventeenth century, this English economist was among the first in his profession to highlight division of labor as a beneficial practice:

...the Gain which is made by Manufactures, will be greater, as the Manufacture it self is greater and better. ...each Manufacture will be divided into as many parts as possible, whereby the Work of each Artisan will be simple and easie; As for Example. In the making of a Watch, If one Man shall make the Wheels, another the Spring, another shall Engrave the Dial-plate, and another shall make the Cases, then the Watch will be better and cheaper, than if the whole Work be put upon any one Man. And we also see that in Towns, and in the Streets of a great Town, where all the inhabitants are almost of one Trade, the Commodity peculiar to those places is made better and cheaper than elsewhere. ... (Petty 1899, Sect. 6)

This intellectual exercise would only increase the castaway's misery and anxiety. He can but realize that he will be materially worse off alone. Society is obviously a complex formation, the different members of which perform separate roles and functions necessary for individual and collective survival and progress. The task of forming a society anew and alone is daunting. Even if he remains with Socrates in

Plato's *Republic*, he will observe that at least a farmer, a builder, a weaver, and a shoemaker are needed; much worse if he entertains needs that go beyond the essentials. Cowper is indeed quick to display Selkirk's misery:

O solitude! where are the charms  
That sages have seen in thy face?  
Better dwell in the midst of alarms  
Than reign in this horrible place.

That a castaway has very little to choose from can be traced to the very limited labor force he has available and the size of the market he has the intention of serving. Both equate to one person. The entire labor force of one will be systematically dedicated to a restricted set of tasks, all required for survival in wilderness. Despite the specialization that may thereby occur, division of labor will not take place. If the castaway is an architect and is intent on building a shelter, imagining and drawing it will not do much: he must do himself all the subsequent tasks needed to complete the shelter, from finding a sturdy tree and gathering materials, to building the frame, walls, and a bed. If, instead, he only builds beds and manages to produce several, then his common sense is failing, or he is planning a second residence, or both.

Division of physical labor cannot develop in a society of one, as both the demand and the supply sides are severely restricted. In any problem he faces, the castaway dictator will have free reign, but only within the very limited set of potential solutions he himself creates, or the natural environment provides independently of his efforts. He is deprived of alternatives that society can make available through production that is organized as William Petty described. Relative to a specialized fellow who can rely on others' discharging their specialized responsibilities, the castaway will not only be deprived of the effects of their physical labor: *the knowledge that they develop through the performance and study of the tasks will also be missing.*

## 2.2 From Division of Labor to Dependence on Advice

"The greatest improvement in the productive powers of labor," wrote Adam Smith (1904) in his *Wealth of Nations*, "and the greater part of the skill, dexterity, and judgment with which it is anywhere directed, or applied, seem to have been the effects of the division of labor." As Petty's watchmaker shows, the division of labor consists of having a worker perform one or a few tasks involved in a production process, instead of having the same person perform many or all tasks in the same process. Each worker can consequently concentrate on some tasks, and thereby gain from not having to switch between tasks and from the improved dexterity in the necessary manipulations. The worker specializes not in the process itself, but the tasks he has been delegated.<sup>1</sup>

---

<sup>1</sup> "This great increase of the quantity of work which, in consequence of the division of labor, the same number of people are capable of performing, is owing to three different circumstances; first,

Specialization of physical labor has its parallels in the specialization of its mental variant. In a pin factory – to borrow Adam Smith’s classical example – it was observed that if one worker only strengthens the wire of a pin and another only points it, more pins will be produced than if each worker individually produced entire pins. Apart from the gains in productivity, skills will differentiate. The worker specialized in a task will know more about that task when compared to a colleague who specializes in another step of the process. In this trivial case, it is through trial and error, and initially the observation of others performing similar duties that the task is learned and perfected. When the task is complex and/or important enough, knowledge of the task deepens. Disciplines and professions develop. Experience is generalized and codified. Those interested in the particular tasks may have access to education, which in turn facilitates their future engagement within professions that can benefit from such preparation. Some of the disciplines will require advanced method of inquiry and impose variously precise rules on the characteristics of the knowledge that is perpetuated.<sup>2</sup>

A process that demands more recent technology and thinking may better illustrate the extent of contemporary divides in intellectual labor. Decoding the deoxyribonucleic acid of an organism – that is, determining the exact order of the individual chemical building blocks, or bases, that make up the DNA – can be very crudely split into five stages: (1) divide long sequences of DNA into fragments of a size appropriate for subsequent analysis; (2) feed the fragments to bacteria to produce millions of copies that act as raw material for subsequent steps; (3) distribute all available fragments to four different solutions, each used to tag a particular genetic letter; (4) pipe tagged fragments to gel-filled tubes to sort fragments according to size; and (5) read the tags on each fragment to obtain a genetic sequence for the corresponding fragment of DNA. Not unlike the watchmakers that William Petty talks of, the engineers and scientists working to decode a genome will be specialists of well-delimited tasks, who coordinate to realize the process together. While Adam Smith’s depiction of pin making seems to sum up most of what there in fact is to pin production, the preceding sketch of genome decoding hides all of the scientific, technological, and industrial advances that were necessary to enable the efficient performance of the process. New types of genetic markers were needed, along with new experimental and computational strategies for cloning large DNA fragments. Little would have been possible without the automation of DNA sequencing. Each of these terms merely labels entire disciplines of science and industry, each involving many specialties. Clearly, the development of the necessary technology and

---

to the increase of dexterity in every particular workman; secondly, to the saving of the time which is commonly lost in passing from one species of work to another; and lastly, to the invention of a great number of machines which facilitate and abridge labor, and enable one man to do the work of many.” (Smith 1904, Book I, Chap. I)

<sup>2</sup> The United States Department of Labor’s *Dictionary of Occupational Titles* is a caricatural illustration of how division of labor continues to advance. This publication’s 1991 edition listed more than 12,000 generic job titles. The dictionary has since been discontinued.

methods required the involvement of highly specialized knowledge from various areas and the coordinated division of both physical and intellectual labor:

It took most centers a while, however, to learn how to organize the most effective teams to tackle a big science project. John Sulston, director of the UK's Sanger Centre (now the Sanger Institute) from 1993 to 2000, recalls that 'at first everyone did everything,' following the tradition of manual sequencing groups. . . . However, it soon became apparent to Sulston and others that, *for the sake of efficiency and accuracy, it was best to recruit staff of varying skills — from sequencing technology to computer analysis — and to allocate the work accordingly.* (Collins et al. 2003, p. 286; emphasis added)

The approach proved successful: the best-equipped laboratories in the mid-1980s could sequence about 1,000 base pairs (i.e., structural units of DNA) a day, while in 2000, sequencing centers could collectively sequence 1,000 base pairs a second (Collins et al. 2003).

While broad parallels can be drawn between a pin factory and a DNA sequencing center, the two workplaces stand in stark contrast in one important respect. Namely, *it takes very different resources to understand the details of the two processes to the level required to match the knowledge held by the specialists of these steps.* In Adam Smith's pin factory, it takes 18 steps to make a pin. Each step involves simple manipulations and tools, the characteristics and use of which can be understood and learned within a reasonable amount of time. It is an understatement to say that it is more difficult to do the same for DNA decoding. Even without focusing on the problems of automation in that process, advanced knowledge of biochemistry, genetics, and molecular biology is needed to understand the chemical and biological mechanisms at play, while statistics and algorithmics are necessary for the treatment and interpretation of the collected data. Control of nuclear fission and exploration of space are other prominent examples of highly complex endeavors, which require the accumulation and use of knowledge of scope and depth beyond the practical reach of an individual. Reaching any such milestone required advanced specialization both of physical and intellectual labor. This is not to say that only such unique undertakings are in their whole beyond an individual's reach: whatever requires advanced science and technology shares this same trait. It is unsurprising that a scientifically and technologically advanced economy would favor the specialist over the generalist, as we can clearly see from the way advanced education is organized.

Recall that the generalist castaway cannot benefit from the division of physical labor. He is unlikely to go far in the division of intellectual labor, for just as his time will be occupied with survival, his knowledge is unlikely to move beyond that relevant to the physical tasks at hand. If we could consider the entire body of knowledge in his primitive economy, that is, all knowledge he has access to, we can see that he will command all of it. In presence of two or more people, division of labor can advance together with the division of intellectual labor, and specialist knowledge will develop. The body of knowledge in the economy with specialization will be comparatively larger than that of the solitary castaway. However, relative to the castaway, the individual in the economy of many specialists will command a

smaller share of total knowledge available in the economy.<sup>3</sup> The specialist will have particular knowledge related and tailored to the performance of one or a limited set of tasks. By performing the task, he may advance that knowledge, possibly to the extent that pushes specialization and division of labors further.

To specialize is to acquire knowledge relevant to the task of interest. Since doing so uses rare resources, there is an opportunity cost: to specialize in one field is to forgo the benefits of doing the same in another. This opportunity cost manifests itself in a clear and visible way in decision situations. Suppose that the castaway is specialized in botany, with a pronounced interest in edible plants. To distinguish those that are edible from others, he draws on what he has previously learned. Relative to a fellow who is not a specialist in the same field, the castaway has access to a body of knowledge that is relevant with regards to the decision situation, and is almost certainly better off in this particular case. If this botanist instead set out to build a boat, and if that was to be the first boat he ever built, then a professional constructor of boats would certainly be better off than the botanist.

To specialize, then, is to acquire knowledge that is most relevant only for some decision situations; in others, the specialist changes his role and reverts to a generalist. Aware that his knowledge is limited in fields other than his own, the individual can only seek shortcuts to the knowledge available elsewhere: (if) aware of own ignorance, he will ask for advice from others. As a recommendation on a decision or course of conduct, advice acts precisely as such a shortcut. In the matter of pin manufacture, consult the pin maker; to build a house or bridge, refer to the architect and to the engineer.

*Division of labor goes together with a dependency on advice.* While the efficient production of pins, watches, but also automobiles and aircrafts requires the division of physical and intellectual labor, efficiency does not arise simply out of the division of labors. Coordination of the specialists is necessary to avoid the waste of the various resources used in such processes, and thereby claim efficiency. A specialist's knowledge will tell him what to do within a process, but it is through coordination that he will know when to do it, how often, whose work he will be building on, and who will in turn be building on his labors. Division of labor creates dependencies

---

<sup>3</sup> "The 'Jack-of-all-trades' is less useful than the specialist in economies with advanced technologies and an extensive human capital base. Although workers in modern economies have considerable knowledge of principles and have access to complicated technologies, a typical worker also commands a very much smaller share of the total knowledge used by the economy than do workers in simpler and more backward economies. [...] An 'expert' has been facetiously defined as 'someone who knows more and more about less and less.' Highly specialized workers are surely experts in what they do, and yet know very little about many other skills found in a complex economy. Modern expertise comes partly at the expense of narrowness, and of ignorance about what other people do. [...] Greater knowledge tends to raise the benefits from specialization, and thus tends to raise the optimal division of labor. This helps explain why workers become more expert on narrower ranges of tasks as knowledge grows and countries progress. Increased specialization in turn raises the benefits from investments in knowledge, so that the growth in tandem of specialization and investments in knowledge may allow an economy to continue to develop." (Becker and Murphy 1992, pp. 1146,1157).

between the specialists, entailing thereby the necessity for coordination. It is further clear that there is no coordination without the exchange of information between the specialists. Regardless of the form that the coordinating information will take – be it spoken, written, involve unspoken observation, or other – its very purpose is to orient others' courses of action. *That division of labor goes together with the dependency on advice is due to the necessity for coordination.* Specialists cannot collectively be efficient without coordinating their efforts, which in turn makes them rely on advice as the information that coordinates. Advice is a tool of coordination.

## 2.3 Autonomy and Coordination

To suggest that advice is a tool of coordination requires that we look into the question of whether advice interacts with the choices of the individual. This is essentially a question of how the presence of advice relates to the autonomy of choice. If advice does not affect individuals' choices, its role for coordination is irrelevant; otherwise, it is proper to claim that coordination happens through advice.

Autonomy refers to the ability to make one's own laws. As the solitary castaway clearly has no one's laws to obey but his own, he benefits from the autonomy of choice. His decision-making is entirely independent of advice.

We called it ironic that a castaway has so much autonomy in choice, but so little to choose from. Both are hard, if impossible to have together. In a society of two or more, advice becomes available to inform choice. As the division of physical and intellectual labor advances, dependence on advice becomes more prominent. In a technologically advanced society, the division of labors and specialization will have progressed and thereby ensured a demand for advice. When in the role of a specialist, an individual can offer advice from personal expertise. When he reverts to the generalist, he may demand advice. To escape the castaway's irony by increasing the range of alternatives in choice thus goes together with another irony: more options seem tied to less autonomy in individual choice. It is not the increase in alternatives itself that does so, but the conditions necessary for such increase to happen. Division of physical and intellectual labor is necessary to benefit from those alternatives, which result from the labor of the community. Division is inseparable from the specialization of individuals. In own areas of expertise, every individual can offer advice. Much of the advice is codified, as a visit to the library illustrates. Solicited or not, it will often be available precisely because it can influence choice. Advertisements are an obvious example. It appears that no nontrivial decision can be taken independently from advice.

To be a tool of coordination, advice should have the capacity of constraining choice. It should eliminate alternatives, occasionally to the extent that only one remains. To accept advice can consequently either simply orient, or fully determine choice. Laboratory experiments on how advice interacts with choice support such intuitions. Economists have studied the role of the so-called naïve advice, which does not come from experts, but is of a word-of-mouth kind, and based

mostly on limited prior experience. In Andrew Schotter's experiments from the 1990s (Schotter 2003; Schotter and Sopher 2006), subjects are engaged in intergenerational ultimatum games. In an ultimatum game, which places players in a bargaining situation, two players need to divide some given sum of money between them. One player called Sender proposes first how to divide the amount. The Sender indicates the part of the amount he wishes to take. The second player called Receiver is then asked to either accept or reject the Sender's proposal. If the Receiver accepts, the amount is split according to the proposal; otherwise, both players receive nothing. In an *intergenerational* ultimatum game, several ultimatum games are played in a sequence: a Sender/Receiver pair plays, then another pair plays, and so on. Each pair plays once, and after each play, the current pair is replaced by the next pair. The next pair can see some or all of the previous plays, that is, what the prior Senders offered, and whether the corresponding Receivers accepted the offers. There is also explicit advice. The Sender of the previous game  $t - 1$  advises the Sender of the current game  $t$  the amount to offer, and gives a brief justification of that advice. The Receiver of  $t - 1$  suggests to the Receiver in  $t$  the minimal acceptable amount. Results of these and similar experiments in laboratory conditions support the arguments offered above:

- Subjects tend to follow the advice they receive, even though this is naïve advice; it is naïve in the sense that the advisors are not more experienced in the matter at hand than the advised
- Advice affects how subjects behave in the experiments, since it is observed that those who receive advice act differently than those who receive none
- If given the choice between advice or the information upon which that advice was produced, subjects tend to take advice
- Giving or receiving advice forces the decision-maker to think about the decision problem in a different way from the way they would have done in absence of advice

Table 2.1 lists additional observations supported by empirical evidence, illustrating that decision-making in presence of advice does differ from what would occur in the absence of advice. Studies which resulted in these findings are typically conducted in settings organized as the so-called judge-advisor system (Bonaccio and Dalal 2006), where the *judge* is the decision-maker who makes the decision after having received advice from the advisor. Advice takes the form of a recommendation in favor of an alternative in the decision problem (e.g., "Choose option X"). In some cases, the advisor is also asked to give to the decision maker an estimate of confidence in the validity of the advice (e.g., "Choose option X; I (the advisor) am 85 percent sure that it is the best option"). Participants enter the laboratory, are assigned a role (either decision-maker or advisor), and told that it is the decision-maker who chooses, so that he can accept or reject the advice that the advisor provides. The advisor and decision-maker both read the information about the decision task. The latter makes an initial decision and, if asked, gives an estimate of his confidence in the appropriateness of his choice. The advisor does not know what this initial decision was. He is asked to make a recommendation and, if asked, give an



**Table 2.1** Findings about advice-giving and advice-taking that are supported by empirical evidence. Adapted from Silvia Bonaccio and Reeshad S. Dalal’s survey of research on advice and decision-making (Bonaccio and Dalal 2006)

Findings about advice-giving and advice-taking
Sharing of accountability for the outcome of a choice and the improvement in the result of a choice motivates the decision-maker to seek out advice (Harvey and Fischer 1997)
New information or alternatives are considered by the decision-maker when he interacts with others before choosing (Schotter 2003)
The framing effect (i.e., when the choice is not independent of how alternatives are presented to the decision-maker) is less pronounced when the decision-maker receives advice from a source he perceives as credible (Druckman 2001)
There is social pressure on the decision-maker not to reject freely offered advice, which if rejected may not be proffered again in the future (Sniezek and Buckley 1995)
Own opinions are more important (i.e., have more weight) to the decision-maker compared to those he receives from an advisor (Gardner and Berry 1995)
Advice is perceived as more helpful and less intrusive when it is offered by an advisor considered as an expert by the decision-maker (Goldsmith and Fitch 1997)
Advisor’s good reputation is gained with difficulty, but lost easily when the quality of advice decreases in the eyes of the decision-maker (Yaniv and Kleinberger 2000)
To estimate the quality of advice, decision-makers use information and explanations about their advisors’ forecasting strategies (Yates et al. 1996)
Decision-makers are responsive to advice coming from those of greater age, education, life experience, and wisdom than the decision-makers (Feng and MacGeorge 2006)
Paid advice is considered as more important by decision-makers than advice received for free (Gino 2006)
Decision-makers tend to discount advisors whose recommendations vary considerably from those of other advisors (Harries et al. 2004)
Recommendations coming from more confident advisors are followed more often than those given by less confident advisors (Phillips 1999)
A decision-maker’s confidence in his choice is higher after the advice was received and the choice made. This may be due to the decision-maker forming a rationalization of their choice based on the advice (Heath and Gonzalez 1995)

expression of confidence in his recommendation. The recommendation is then given to the decision-maker, who makes the final decision. In contrast to experiments in which naïve advice is used, studies using variants of the judge-advisor system give richer findings, which in many cases confirm what seems intuitive, e.g., that advice from experts is deemed less intrusive and more helpful, that the decision-maker



considers his own opinions as more important than those of the advisors, that expert reputation is gained with difficulty but lost easily.

Beyond controlled environments, we can rather straightforwardly observe that even the decision-makers who seemingly can exercise high autonomy of choice will not necessarily do so, and instead rely on advice. Dictators are particularly illustrative: by being an individualistic ruler, any dictator of many intends to choose alone and abhors external influence. The castaway dictator can behave in such a manner, for there is effectively no other way. Any other dictator, however, must rely on advice because of the complexity of the effort needed to remain in place. To do so, the dictatorship must be consistent continually and in its various sources of power. It will need to maintain the authority that gives legitimacy to the regime, ensure access to skills, knowledge, and material resources it needs, enforce sanctions and communication that acts on the psychological attitudes of its subjects (Sharp 1973). Same applies for business executives. Even an authoritarian and individualistic ruler cannot choose independently of advice. Since remaining in power requires various specialized knowledge and skills, and as there is an opportunity cost to specialization, it is not surprising to see, as discussed in the next section, that dictatorships must rely on advice of many specialists.

## 2.4 Coordination Through Advice

Advice is a tool of coordination in the sense that it allows the transfer of knowledge between specialties, whereby the aim of the transfer is to inform decisions. To recognize this is to observe that advice is a manifestation of coordination. This is not to say that all coordination can be reduced to the exchange of advice (i.e., that advice is the *only* manifestation of coordination), but only that different ways of coordinating are likely to be accompanied by advice of various form and content, while ways in which advice is exchanged may also differ.

Once we admit that division of physical and intellectual labor goes with coordination, and that coordination can happen through advice, a different perspective of coordination mechanisms becomes of interest. Namely, we can ask what characteristics advice has – in terms of form, content, and advice-giving processes – within different means of coordination. As we shall see below, much of economics, management, politics, and law are concerned with coordination. Given the degree to which these spheres of interest in effect study and promote ways with which civilization deals or intends to deal with coordination problems, it will be argued that the study of advice positions itself as a fundamental issue of interdisciplinary relevance. To the extent that political, economic, managerial, legal, and religious models can be understood as systems of advice combined with mechanisms that enforce consent to advice, we shall see that the issue of how an individual assesses advice effectively becomes a determinant of his relationship to and position within coordination mechanisms and thereby within the political, economic, legal, and other such realms.

## 2.5 Advice in Political Coordination

In promoting with almost religious zeal a form of early communism, an 1894 political pamphlet in Britain, called *The Labour Leader* (Muse 2009) announced the collapse of party politics. For the price of one penny, the interested reader would open the pamphlet to find a series of quotes under a dominating title “What leading thinkers have said,” followed by a more elaborate argument by the writer. Among the great thoughts, a politician was quoted saying that “party spirit makes people abjure independent thinking. Some range themselves on one side, and some on the other, as they used to do in their school games, and with about as much reflection.” It remains unknown if this was sarcasm. The benefit of hindsight tells us that such collapse has been announced much too early. Political science from the 1930s to the 1950s recognized party government to be a distinguishing feature of then-modern politics. Called a revolutionary change in the conduct of public affairs, the party system was contrasted at the time to hereditary authority, continual resort to violence, and religious symbolism as the sources of political power (MacLeod 1931; Clokie 1949; Leiserson 1957).

The role of coordination that political parties perform is immediately apparent when we consider their purpose. A party organizes those interested in acquiring and exercising political power toward the achievement of specific objectives. They thereby coordinate the efforts of individuals toward the acquisition of political power, and subsequently their attempts at the realization of various aims. In performing that role, political parties can be said to constitute in contemporary societies an obvious link between those within and outside the formal institutions of government. In multi-party political systems, where access to power passes through the voice of the electorate, a party will need to coordinate internally not only the acts of its members, but also of the individuals whose votes they need. What we consequently systematically see them doing before elections is advocacy through communication with the electorate. Parties proceed by laying out the party platform, devised to distill and communicate a party’s world view, offering thereby opinions on the general courses of action that they favor. This is an old tradition, as we can see from *The American Presidency Project* at the University of California, which archives the platforms of US political parties. The US Democratic Party Platform of 1840 states, among others that “. . .the separation of the moneys of the government from banking institutions, is indispensable for the safety of the funds of the government, and the rights of the people” (Anonymous 2009). In addition to the advice on the separation of state-owned financial institutions and private banks, it also recommends equal treatment of citizens, the governmental collection of revenue only to the extent needed to cover the necessary expenses of the government, and so on. While the form of such advice has changed today to a great extent – the democrats’ platform of 2008 speaks in a different terminology and tone, and is much longer than the short tract of the 1840s – the purpose has remained essentially the same: summarize the message to the electorate, ensure the uniformity and the consistency of that message, and use it to coordinate with those within and outside the party. While the message does directly advise on governance, it only indirectly advises

the voters on whom to tick at the polls. If the individual favorably evaluates that advice, and has no substantive disagreements with the worldview it advises, he is expected to align his vote with the party in question. Coordination of this sort could not happen if advice was not made explicit.

That some considerations found in modern party programmes may have been advocated for so long that they are no longer recognized simply as advice does not change the fact that they do amount simply to recommendations on courses of action.

The exercise of political power requires consent to advice, that is, acting as advised. Since any political programme is but advice on governance, coordination will not happen without seeking the consent to given advice. Different political systems seek consent to advice in a diverse ways. Such variations illustrate the differences in the recommendations given on the acceptance of advice itself. How consent is sought within a political system thus illustrates, beyond the advice of the party programme, the advice of more general content and purpose: that which advises how to accept any other advice that the government may dispense. Beyond the very content of the party programme, and occasionally within the party programme itself, we can find observations that pertain not directly to the concerns of economic, internal, or foreign policy, but those of a more abstract nature. In the 2008, US Democratic Party Platform, and under the heading “Build Democratic Institutions,” it is said that the party “will increase . . . support for strong legislatures, independent judiciaries, free press, vibrant civil society, honest police forces, religious freedom, equality for women and minorities, and the rule of law.” As we shall illustrate below – in briefly discussing the status of advice in dictatorship and in democracy – the government will not only advise on what the tax rate should be, and what countries will be considered friends or foes, but also the degree to which the advice dispensed by the government can be questioned. In effect, it will advise the extent to which the independent assessment of advice is (un)welcome.

### **2.5.1 Dictatorship and Consent**

At the end of the 1920s and at the beginning of the 1930s, Germany was a multi-party political system. The Nationalist Socialist German Workers Party of the 1930s was one of more than a dozen political parties vying for attention and acceptance of voters in parliamentary elections. From the perspective of a critical nonmember, their ideas could only amount to opinions, dispensed as recommendations on how the government should be run. Whether to accept this advice, and thereby orient own decision-making accordingly could, or at least should have been decided individually, and in the manner of what Herbert Marcuse calls the “last analysis” in his *One-Dimensional Man*: the individual should decide only if free to take either option. The parliamentary elections of 5 March 1933 resulted in the formation of a coalition of the National Socialists and the German National People’s Party, holding 52.5% of seats of the parliament. In this respect, Karl Loewenstein,

a scholar of constitutional law, notes that “the elections from which the National Socialists derived the legal claim to revolutionize the fundamental order of the state yielded only a very slender margin of majority which as it is known was immensely exploited by the victorious parties. . . The constitutional basis on which the reconstruction of Germany on National Socialist lines rests was created mainly by fraud and terrorization.” (Loewenstein 1937, pp. 540–541). That the acceptance of advice, in this case for whom to vote, was sought with violence before the elections was a signal for subsequent events. The freedom to publicly question the advice given was eliminated in a rapid and systematic manner through the reform of the constitution. The principle of equality before law was abolished, along with the *nulla poena sine lege* principle, which guarantees that one cannot be punished for acts that are not prohibited by law. Political parties, freedom of assembly, of association, and of public opinion were all prohibited; violations of the right to private property went unnoticed. In such extreme cases, law becomes merely the rewriting of leaders’ advice. Aware that their opinions may change, the leaders choose to make law by the very act of dispensing advice:

The new law as the binding expression of the Leader’s will claims precedence of right over all other rules of law. . . Concerning the rules passed prior to the National Socialist revolution [any] judge is bound only by his conception of the National Socialist ‘spirit’. . . This implies that parts of the Weimar Constitution continued to be in force although derived of their formal character and subject to being amended at any time by governmental decree, government ordinance and statutes passed by the Reichstag. . . The essence of the political revolution in Germany may be expressed adequately by the simple statement that the will of one man alone, the Führer, is sovereign, free from any constitutional limitations whatsoever. (Loewenstein 1937, pp. 546, 554)

The freedom to oppose what is quite simply someone’s advice thereby disappeared. Consent is obtained making alternatives illegal and their choice sanctioned. Beyond a mechanistic consent to advice from fear of oppression, the dictator expects the citizen to accept the world view, that is, the ideology, that advice reflects. The framework of law in a dictatorship advises the individual to suspend independent thinking.

### 2.5.2 Under Democracy

Dictatorship is but one of various forms of governance that requires the individual to suspend the independent assessment of advice. If we are interested in a more general picture, it can safely be argued that such suspension is a trait of what Karl Popper calls tribal societies in *The Open Society and Its Enemies*. These are groups of people sharing a strong cultural or ethnic identity, and beliefs that regularities observed in both nature and social life are due to something supernatural. Despite the many differences that may distinguish one tribe from another (what social structure they may promote, how they organize production, how they perform the transfer of power, and so on), what they share is that individual decision-making happens

within the rigid bounds of such belief. Shared beliefs have a clear value for the coordination of individual action, for they make the individual's behavior predictable when he is called to interact with others in both private and public matters. The beliefs will either directly prescribe or indirectly guide the behavior in as different situations as those in which the individual is called to act in some way by a figure of formal authority, or how he perceives and protects the results of his labor. This is not to say that the beliefs cannot change, but that when they do, they are not based on independent thinking about the prior practices.

Just as the dictator will seek to impose the advised course of action by rendering any other illegal, so will the tribal chief or king support the advised action through myths and taboos. In both cases, it is fear from some form of anticipated violence that will deter choice that goes against or past advice. Religions tend to ask that independent reflection on the relevance or appropriateness of advice be suspended. Psalm 119:105 from the King James' Christian *Bible* conveniently summarizes the attitude that a subject is to have of advice dispensed by the various texts: "Thy word is a lamp unto my feet, and a light unto my path." To complete this picture, note that the *Catechism of the Catholic Church* suggests that the scriptures "govern the whole Christian life" (Catholic Church 1992). Bob Dylan nicely puts the role of anticipated violence on choice in the song *Highway 61 Revisited*, where he caricatures an infamous episode from the Old Testament:

Oh God said to Abraham, "Kill me a son"  
 Abe says, "Man, you must be puttin' me on"  
 God say, "No." Abe say, "What?"  
 God say, "You can do what you want Abe, but  
 The next time you see me comin' you better run"  
 Well Abe says, "Where do you want this killin' done?"

Advice that comes with incentives or sanctions is not peculiar to dictatorships or Christianity. The economist used to capitalism will be surprised by the extent to which Islam meddles into a banker's affairs. Sūratu al-Baqarah, the second chapter of the *Qur'an*, advises: "...relinquish what remains [due] from usury, if you are believers." It then clarifies how this advice should be read: "But if you do [it] not, then be apprised of war from Allah and His messenger" (cited in Labib 1969). In Machiavelli's words, "it is easy to persuade [people] of something, but difficult to maintain them in their belief; and therefore it is essential to arrange things in such a way that, when they no longer believe, they can be forced to believe" (Machiavelli 2009, Sect. 6).

Despite the taboos of current societies, Popper argues that the important difference from tribal societies is the widening field of personal decisions. When individual reflection and questioning is tolerated, courses of action alternative to those prescribed through taboos can be entertained. This transformation from the tribal, or closed, to the society open to critical attitude is in Popper's opinion "one of the deepest revolutions through which mankind has passed" (Popper 2002, p. 188). Given the persistence in the twentieth century, through dictatorships among others, of the mechanisms deployed to ensure consent to advice, it is more appropriate to say that mankind has not passed this transformation, but that the process has already been going on for at least two and a half millennia.

The tension between the societies more or less closed to personal decision, and thereby the independent assessment of advice has in a sense been officially announced at the outset of the Peloponnesian War, waged from 431 to 404 BC and which opposed Athens to the Sparta-led Peloponnesian League. At the public funeral of the first who fell for Athens, the leader of the city-state delivers a speech that Thucydides recounts in one of the earliest scholarly works of history, the *History of the Peloponnesian War*. Pericles not only honors the dead, but identifies a number of ideas that he sees as confronting Athens to its opponents:

Our constitution does not copy the laws of neighboring states; we are rather a pattern to others than imitators ourselves. Its administration favors the many instead of the few; this is why it is called a democracy. If we look to the laws, they afford equal justice to all in their private differences; if no social standing, advancement in public life falls to reputation for capacity, class considerations not being allowed to interfere with merit; nor again does poverty bar the way, if a man is able to serve the state, he is not hindered by the obscurity of his condition. The freedom which we enjoy in our government extends also to our ordinary life. There, far from exercising a jealous surveillance over each other, we do not feel called upon to be angry with our neighbor for doing what he likes, or even to indulge in those injurious looks which cannot fail to be offensive, although they inflict no positive penalty. But all this ease in our private relations does not make us lawless as citizens. Against this fear is our chief safeguard, teaching us to obey the magistrates and the laws, particularly such as regard the protection of the injured, whether they are actually on the statute book, or belong to that code which, although unwritten, yet cannot be broken without acknowledged disgrace. . .

If we turn to our military policy, there also we differ from our antagonists. We throw open our city to the world, and never by alien acts exclude foreigners from any opportunity of learning or observing, although the eyes of an enemy may occasionally profit by our liberality; trusting less in system and policy than to the native spirit of our citizens; while in education, where our rivals from their very cradles by a painful discipline seek after manliness, at Athens we live exactly as we please, and yet are just as ready to encounter every legitimate danger. . .

Nor are these the only points in which our city is worthy of admiration. We cultivate refinement without extravagance and knowledge without effeminacy; wealth we employ more for use than for show, and place the real disgrace of poverty not in owning to the fact but in declining the struggle against it. Our public men have, besides politics, their private affairs to attend to, and *our ordinary citizens, though occupied with the pursuits of industry, are still fair judges of public matters; for, unlike any other nation, regarding him who takes no part in these duties not as unambitious but as useless, we Athenians are able to judge at all events if we cannot originate, and, instead of looking on discussion as a stumbling-block in the way of action, we think it an indispensable preliminary to any wise action at all. . .* And it is only the Athenians, who, fearless of consequences, confer their benefits not from calculations of expediency, but in the confidence of liberality.

In short, I say that as a city we are the school of Hellas, while I doubt if the world can produce a man who, where he has only himself to depend upon, is equal to so many emergencies, and graced by so happy a versatility, as the Athenian. . .the admiration of the present and succeeding ages will be ours, since we have not left our power without witness, but have shown it by mighty proofs; and far from needing a Homer for our panegyrist, or other of his craft whose verses might charm for the moment only for the impression which they gave to melt at the touch of fact, we have forced every sea and land to be the highway of our daring, and everywhere, whether for evil or for good, have left imperishable monuments behind us. (Thucydides 2004, paragraphs 37–41; emphasis added)

It would not be displaced to hear a similar speech today on any, even remotely democratic political rally, for it espouses the ideas of participation in governance (“administration favors the many instead of the few”), equality before law (“they afford equal justice to all in their private differences”), individual competence (“advancement in public life falls to reputation for capacity”), personal liberties (“we do not feel called upon to be angry with our neighbor for doing what he likes”), rule of law (“teaching us to obey the magistrates and the laws”), openness to foreign citizens (“never by alien acts exclude foreigners from any opportunity of learning or observing”), regard for various opinions (“we think it [discussion] an indispensable preliminary to any wise action”), and progress informed by learned facts (“far from needing a Homer for our panegyrist. . . we have forced every sea and land to be the highway of our daring”). Although certainly an impressive address, it is important to recall some aspects of its context, beyond the need to inspire a people assembled in mourning. As far as equality before law is concerned, it is enough to note that Athens admitted slavery. To paraphrase a famous line from George Orwell’s allegorical *Animal Farm*, it seems that people of Athens are equal, but some are more equal than others. Independent thinking, though famously proclaimed in the funeral oration, was selective in Athens in addition to being biased by tradition and taboo. With regards to taboos, Pericles advises widows as follows, before closing the address:

. . . if I must say anything on the subject of female excellence to those of you who will now be in widowhood, it will be all comprised in this brief exhortation. Great will be your glory in not falling short of your natural character; and greatest will be hers who is least talked of among the men, whether for good or for bad. (Thucydides 2004, paragraph 46)

Despite the important nuances that should be kept in mind when reading Pericles’ funeral oration, it remains a strong example of an attitude toward advice that stands in stark contrast to that in any form of authoritarian rule. Particularly important in this respect is the part emphasized in the longer quote above. It can in effect be read as pertaining to advice that the government asks from its citizens under a democratic regime: “we Athenians are able to judge at all events if we cannot originate.” In another translation, this same idea is stated more tellingly: “although only a few may originate a policy, we are all able to judge it” (cited in Popper 2002, p. 199). Pericles not only gives advice in his speech on what Athens stands for, but also applauds the advice that citizens can provide *and* the very idea that they are allowed to advise, or in other words, that they can evaluate the advice that is offered to them by government. In the Athenian democracy, the government both dispenses advice and is open to recommendations from the citizens who are usually not engaged in daily politics.

Direct democracy, as preached in Athens of the fifth century BC, has a modern variant that coexists with the various representative democracies. While an elected minority exercises political power in representative democracies, the direct form operates differently. In modern USA, for instance, citizen’s advice can be solicited on state and local (but not the national) levels via initiative, referendum, and recall: initiative serves to propose a legislative measure or constitutional amendment; referendum asks if the voters are in favor of an existing piece of legislation; and



recall allows the public to ask for a vote on the continued tenure of an official. Such tools coexist with the representative system to allow the expression of voter sentiment when the representative process is dysfunctional. They put the voter in the position of the advice-giver and ensure to some extent that the voter's advice is followed. This is not to say that these tools are flawless, or even any good at all. Thomas Cronin, a political scientist, observes in his *Direct Democracy* that a number of recurring problems persist, such as questionable signature-gathering, confusing language in proposals put to vote, and the role of advertising techniques in the promotion of proposals. Finances also seem to play an important role, for "in about 20 percent of cases, the underdog or more weakly financed side wins" (Cronin 1989, p. 113). Nevertheless, what is usually vaguely understood as a democratic form of government has a significantly different attitude toward voters' advice than authoritarian rule.

### 2.5.3 *Advice and Meta-Advice in Politics*

Reading Pericles, and having noted some traits of the dictatorship in Germany of the 1930s and 1940s, it would appear that the attitude of the government toward the independent assessment of advice is very different in the two cases. In simplistic terms, it seems that the dictatorship receives none from, but only dispenses advice to the people, while a democracy more or less directly both gives to and asks for advice from the people. This has been nuanced, for it was also noted that any but the castaway dictator can remain in power without relying on advice in various matters, and that democracy selects the advice it asks and accepts.

Just as the dictatorship makes a precise recommendation regarding the practice of the independent assessment of advice, so does the (modern) democracy. In both of these forms of government, such recommendations themselves are suggestions on courses of action, that is, advice. This is a simple, yet important observation, since it supports the notion that how an individual assesses advice that positions him in relation to political ideas, some of which may have been advocated so intensely that they are taken for granted and *not* seen simply as advice, which the individual can choose himself to accept, reject, or discuss. Simple examples illustrate this: it appears self-evident today that a country needs a written constitution, yet the UK has no such document. It also appears obvious that there must be some formal right to private property for any sort of economic exchange to take place, yet it is only in 2004 that China reformed its constitution to reintroduce the provision that property right is not to be violated, after abolishing the notion of private property from its constitution in the 1950s. As a third brief example, note that for at least several decades before and after the Second World War, many considered it clear that any form of collectivist economy will lead to authoritarian rule, as for instance Friedrich Hayek argued in *The Road to Serfdom* (Hayek 2001), and that individual freedom cannot exist without the free market economy, as Milton Friedman suggested in *Capitalism and Freedom* (Friedman 2002); however, in 2009, the role of the free



market is openly questioned following the global economic and financial crisis that manifested itself after 2007. It seems that the worst in the 2007 and onwards crisis was avoided precisely through the collectivist intervention of big government that has been so troublesome to many free market economists. In all of these cases, we are dealing with recommendations on some or other aspect of private and public life, and thereby are confronted to the question of whether to accept, reject, or discuss advice, and more generally of how to assess advice.

Regardless of the particular form of democracy we may be interested in, any such form will be a particular case of what Robert Dahl, a political scientist, called a polyarchy (Dahl 1971). For a polyarchy to exist, the following institutional requirements should be present: (1) the constitution gives the control over governmental decisions to elected officials; (2) all elected officials are chosen and peacefully removed from office in relatively frequent, fair, and free elections, in which coercion is limited; (3) almost all adult citizens have the right to vote in such elections; (4) almost all adult citizens have the right to run for public offices, the occupation of which is decided at the elections; (5) all citizens have an actually enforced right to freedom of expression, and in particular political expression, which includes criticism of the officials, of the government's decisions regardless of the realm of such decisions; (6) citizens have access to sources of information that are not monopolized by the government; and finally, (7) all citizens have an actually enforced right to form and join autonomous associations, including political organizations (e.g., political parties and interest groups). Modern democratic governance cannot be established without mechanisms, which are intended to support the communication of the results, stemming from the assessment of advice dispensed by the government. The fifth requirement explicitly states this: namely, it opens the possibility for public criticism of the government's choices, and thereby of any advice that the government may choose to promote. Pericles' conception that all can judge policy that only some may originate is thus clearly expressed in the understanding of polyarchy and is very different from the opposing recommendation we have seen as being offered and enforced under authoritarian rule. Of course, there is – as briefly hinted above with regards to the influence of finances through communication on the opinions of the electorate – some non-negligible distance between the utopian definition and the actual use of the instruments of polyarchy to their original purpose.

The attitude a government may adopt toward advice is itself something that starts off as advice. The attitude to advice and to the independent assessment of advice may or may not appear in a party programme, but its limited visibility does not make it inexistent. In a dictatorship then, the very first advice that a subject is given is not some concrete consideration on some specific social, economic, or military issue. Instead, the primary advice that is dispensed is that one should suspend the independent assessment of all advice given by the government. We will call this *meta-advice*: any advice that itself speaks about advice. If then, we accept that the political system has a role of coordination, and that it must dispense advice to realize coordination, the political system will necessarily proclaim meta-advice, which paves the way for all recommendations that may follow. Where polyarchy seems to differ from dictatorship is that it aims to institute mechanisms for the evaluation

of advice. In proclaiming that all can judge policies, Pericles conveys meta-advice: namely, that advice given and received can be discussed before action is taken.

It is evidently different to assess advice that suggests, say, some level of taxation as opposed to another level of taxation, from the advice that suggests the practice of taxation itself as opposed to no taxation at all. Consider the consequences of such assessment on coordination of citizens concerned by that advice – i.e., those who will pay the chosen level of taxes in the first case, and those who either will or will not pay the taxes in the second case. In either case, a decision must be taken before we can say that there is any coordination with regards to the payment of taxes. While the assessment is taking place, and thus no decision is taken, coordination of the citizens on this matter will not happen. It can then be argued that (re)assessment of the practice of taxation will have to be done less often than the assessment of the actual level of taxation, or else coordination will be rather inefficient. Although this efficiency argument has its merits, it should be remembered that looking for efficiency before all other considerations will highlight dictatorship as the appropriate form of government: by leaving public choices to a single individual is undoubtedly more efficient (but only that) than any other form of collective choice.

Meta-advice in a codified form can be found prominently in the United Nations' *Universal Declaration of Human Rights*, in which Article 19 states: "Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers." In saying that one is free to impart information and ideas, it advocates the freedom to give and receive advice, and consequently the freedom to perform the assessment of advice.

That a document formalizes the right to perform the independent assessment of advice should not be confused with the individual's liberty to perform such an assessment. The act of acknowledging the field of personal decision is separate from the existence of that field: the field is there regardless of the public acknowledgment. It is there, however, only to the extent that the individual himself recognizes the ability to assess options and choose. Official recognition can only create more or less favorable conditions for the public expression of personal decisions. The right is needed to avoid sanction when publicly performing, or announcing the result of one's assessment of advice. Settings where such right is absent, and in which the observed actions required the individuals to independently assess recommendations can illustrate that a formally recognized right is not a precondition for such assessment to happen.

Following the period of slavery in the USA, a number of laws that mandated racial segregation were enacted between 1876 and 1965. This established a form of domination that maintained political, social, and economic oppression well into the twentieth century. The social oppression placed the American Blacks very clearly in a subordinate position, as they were effectively living in a separate society, that is, used separate schools, sat at designated places in buses and trains, used separate toilets, had no access to hotels, and so on. The survey data as late as in 1940s suggested that racist attitudes were widely accepted, and that the segregation laws were supported by the white population of the USA (Bobo 1997); these

laws effectively promoted the notions of intellectual and cultural inferiority. The response to such measures has been continuing critique that culminated in the civil rights movements of the 1960s (Morris 1999). The civil rights movement is a result of the long-term assessment of and response to the segregation laws and racial attitudes within the communities of the oppressed. Some explicit forms of critique were pamphlets in the nineteenth century USA [e.g., (Crockett 2001)] and the Harlem Renaissance of the 1920s, which produced protest literature. Various radical ideologies formed both in that period and during slavery and have inspired subsequent boycott movements and protests. Illustrative in this respect is the March on Washington Movement of the early 1940s, which was effective without ever having been held. The intended target of the march to the White House was racial segregation through selective hiring in the US defense industry at the time. Suspecting the march and the embarrass that it would cause to the US who was then engaging in a fight against racist attitudes of the Nazis, Franklin Roosevelt, the president at the time, issued an order to ban racial discrimination in the defense industry. Such acts of non-violent action, which arise out of active criticism of government policies, illustrate the simple but significant observation that Gene Sharp argued for in *The Politics of Nonviolent Action*: regardless of what the government may proclaim or its members may believe to be the sources of its power, this power cannot be exercised without the consent of the people. Regardless of the oppression, criticism exists within the field of personal decision and progressively moves out of it to manifest itself, among others, through some form of violent or nonviolent struggle.

While the example of racial intolerance may seem distant to a citizen of a modern European or North American democracy, this perception does not mean that meta-advice regarding the independent assessment of advice – of which Pericles proclaimed a preliminary form – is dispensed or acted upon equally throughout the contemporary world. Those inclined to quantitative assessments can consider some available measurements of democracy, which seek to provide an aggregate and obviously rough (and thereby necessarily misleading to some extent) picture of the “degree of” democracy in a country. Since these indicators intend to characterize the score of a country along at least all of the institutional requirements for a polyarchy, we can use them here as a proxy for the degree, to which the independent assessment of advice is promoted. *The Economist*, a British weekly publishes each year an “Index of Democracy” based on the idea that “free and fair elections and civil liberties are necessary conditions for democracy, but they are unlikely to be sufficient for a full and consolidated democracy if unaccompanied by transparent and at least minimally efficient government, sufficient political participation and a supportive democratic political culture.” Its 2008 edition surveyed 167 countries (and excluded micro-states), classifying 30 of these as “full democracies,” 50 as “flawed democracies,” while 36 countries are considered as “hybrid systems” and 51 fall under the term “authoritarian regimes.” This places 14.4% of the world population under various “full” democracies, but more tellingly for the present discussion, it leaves 50.1% of the world’s population under some form of authoritarian rule (Anonymous 2008). The World Bank’s Worldwide Governance Indicators serve as aggregate measures of six characteristics of governance and are computed for 212 countries. As we

have noted that the meta-advice (on the open assessment of advice) associated with democratic governance goes together with freedom of speech, the so-called voice and accountability indicator is of interest, “measuring perceptions of the extent to which a country’s citizens are able to participate in selecting their government, as well as freedom of expression, freedom of association, and a free media” (Kaufmann et al. 2008, p. 7). For 2007, Scandinavia, North America, and all European countries west of Poland, along with Japan and South Korea, Australia and New Zealand score high, with significantly lower scores in most of Asia, Africa, and Central and South America. While some divergences are present between The Economist’s and World Bank’s indicators, and though we can certainly argue against the possibility of a precise assessment of a “level of democracy,” it is difficult to argue against their overall conclusions: neither paints an optimistic picture.

A distinction should be made between advice and meta-advice. Both – by being kinds of advice – lend themselves to independent assessment by the individual. The assessment of meta-advice is particularly important because of its relationship to concrete recommendations. The contrast between the use of advice in dictatorship and polyarchy illustrated that meta-advice is not an esoteric notion, but manifests itself either implicitly through acts of government (as in the limitation freedom of expression or assembly under dictatorship) or explicitly via legal documents, some of which are formally recognized worldwide (as in the United Nations’ Universal Declaration of Human Rights). Such recognition need not go with actual application, as was illustrated by the example of the movement against racial segregation and the data on democracy scores. From there on, we looked at forms of governance as means of coordination, and consequently distinguished them on the basis of what advice and meta-advice they use as a tool for coordination.

*What is, then, the link between the observations above and the division of physical and intellectual labor and coordination?* The specialization that goes together with the division of labors creates both an offer and a demand for advice, as we have seen that any specialist will revert to the generalist in any decision situation beyond his own specialty. Moreover, we have argued that advice will serve as a tool for the coordination of the specialists. In a society with advanced division of labors and specialization, the coordination of specialists is essential for further progress. Any form of government in such a society will consequently be obliged to dispense both advice and meta-advice to the aim of coordination. We come to observe then a contrast between dictatorship and polyarchy that may not be immediately apparent. Dictatorship limits the number of situations that call a specialist to revert to the generalist. It does so by placing consequential public decisions in the hands of the dictator. In contrast, any form of governance that admits public and independent assessment of advice obliges the specialist to revert to the generalist whenever he assesses the advice outside his specialty. Polyarchy asks the specialist to revert to the generalist at least every time there are elections, or more generally, every time he is asked to voice his opinion in favor or against some policies. We noted at the outset of this chapter that the botanist castaway is better off in finding edible plants on an uninhabited island than the castaway who is instead specialized in ship building.

A polyarchy may well ask the individual specialized in botany to decide on some problem in ship building.

In mid-2005, the French held a referendum on the Constitutional Treaty of the European Union, in which almost 55% of the electorate chose rejection, 45% hoped for adoption, and the turnout was at about 70% (Anonymous May 30, 2005). The purpose of the treaty was apparently to facilitate the decision-making in the institutions of the EU after the enlargement from 14 to 27 member states. This seemed necessary – e.g., data showed that the number of laws passed every month dropped after the enlargement, whereby the laws being passed were mostly technocratic.<sup>4</sup> There were among the voters both botanists and ship builders, along with butchers, bakers, doctors, and so on, who before the referendum and at the polls were obliged to revert from their specialty to say whether they agree with procedures aimed to improve the decision-making of, for instance, the 785 members of the European Parliament. An overwhelming majority of the voters are obliged to rely on advice dispensed during the campaigning, as it is hard to imagine that the specialist knowledge of edible plants, ship building, meat cutting, bread baking, or surgery will be informative with respect to the choice procedures for an international governing body. It is important not to read this comparison of polyarchy and dictatorship as concluding that dictatorship is somehow more appropriate a form of governance because it may revert the specialist to the generalist comparatively less often than the polyarchy. Horrors of dictatorship are well known and need not be repeated much here; as far as efficiency is concerned, there is no observed regularity saying that faster, rather than slower decisions procedures go together with better outcomes. Instead, we see that once a scientifically and technologically advanced society accepts some form of polyarchy, it will make its specialists dependent on advice almost any time they are called to exercise their rights of participation in governance. Whenever the meta-advice of government advocates the independent assessment of advice by its citizens – as it is in any variant of polyarchy – the way that the individual assesses advice and meta-advice before choosing will not only be an important private matter, but will concern his fellow citizens as well. The individual's personal liberty to assess advice, which may in addition be recognized through meta-advice within a form of government, raises not the question of whether the individual will analyze advice, but rather how he will do so. Finally, if some friction is observed between the visible acts of government and the stated meta-advice of polyarchy (say, freedom of expression), its source is at least in part to be found in how individuals themselves assess advice, and in particular what they do with meta-advice: how they have interpreted it within the context of their own field of personal decision, and how it subsequently affects their actions.

---

<sup>4</sup> The last three directives adopted in 2005 were: "Council Directive 2005/92/EC of 12 December 2005 amending Directive 77/388/EEC with regard to the length of time during which the minimum standard rate of VAT is to be applied.," "Council Directive 2005/94/EC of 20 December 2005 on Community measures for the control of avian influenza and repealing Directive 92/40/EEC," and "Council Directive 2005/93/EC of 21 December 2005 amending Directive 69/169/EEC as regards the temporary quantitative restriction on beer imports into Finland." (Baldwin and Widgrén 2007)

The practice of censorship illustrates that people assess (meta-)advice in various ways. To do censorship is to suppress any form of communication that some individual or group of individuals (i.e., censors) consider unsuitable for a public, whose informational intake these censors set out to control. It seems a rather straightforward affair, where the censor would start from some manual recommending what is allowed and what is forbidden, then proceed to strike out or destroy whatever fits the criteria for rejection. In practice, censorship does start from advice that is dispensed by the government. For instance, the ways the Soviet rule of the 1930s censored arts was particularly effective. Artists were advised to depict heroes and the ideals of communism with convincing optimism, and the official stance was that this line of artistic exploration would provide, “exceptional prospects for manifesting creative initiative, of a choice of diverse forms, styles, and genres” (cited in Wallach 1991, p. 76). While this left the artists guessing what fits the government’s taste, censors were trained to recognize whatever may be questionable to the ideology (Plamper 2001). This was not only based on lists of undesirable works, but also involved the censor interpreting a piece and deciding if it may have an undesirable interpretation beyond the one that first comes to mind. To do so in the case of newspapers, censors were also expected to hold the pages against the light to check for undesirable juxtapositions of text or images: e.g., an entire circulation of brochures of a Stalin speech were confiscated, because “the graphics of the cover were executed in such a way that, when turning the picture, above the columns a Tsarist crown becomes visible” (cited in Plamper 2001, p. 537). In 1930s Russia, censorship advice was explicit and effective, as it was followed publicly by the majority of artists and certainly by the censors of the state (Wallach 1991; Plamper 2001); the advice that was dispensed did translate into intended actions.

While 1930s Russian censorship fits the usual understanding of such matters under an authoritarian regime, where meta-advice says that all council of the government is unquestionable, cases that counter the usual view suggest that (meta-)advice certainly need not be followed as it was originally intended, and that much remains in the assessment to the individual himself. It is since the end of the thirteenth century that publishers in France were subjected to the authority of the university and books were subjected to review by the academia before publication. As Robert Darnton, a cultural historian, suggests (Darnton 1995), this was a practice of censorship, but not a stereotypical one. The review expected of the professors was followed, if favorable, with the approbation of the king, by which he grants the exclusive right of reproduction to the author of the book. Just as the censor could reject a book, he could admire its style and content; censorship letters from the mid-eighteenth century contain comments on style or ideological matters, along with the censor’s reasoning behind such observations. While this does seem mild, underground trade in unreviewed books did exist and views that counter the king and church could have been confined to the audience of the gallows or the Bastille. Still, while they certainly were advised to object to anything inadmissible to church and king, in the analogous way to the Stalinist censors later on, that advice was loosely followed. Darnton mentions scandals, such as when a censor endorsed a translation of the *Qur’an*, after observing that it contained “nothing contrary to the Christian

religion” (cited in Darnton 1995, p. 46). Kings sought consent through violence, as when a sixteenth century edict required anyone publishing a book be hanged, while another in 1757 dictated death to any author of irreligious and incendiary works. “In practice” observes Darnton, “[the censorship system] became increasingly flexible, thanks to the enlightened administrators who bent the rules and, by doing so, created enough room in an archaic structure to accommodate a great deal of modern literature — at least until it all came crashing down in 1789” (Darnton 1995, p. 47). Yet another kind of censorship is present in modern mainstream media. Polyarchy advises media to be independent, in the sense that it should commit to discover and report truth, without being concerned for the harm or gain that this could have for government, enterprise, or individuals. Edward Herman and Noam Chomsky’s *Manufacturing Consent* presents a number cases of self-censorship by reporters and commentators, who “do similar thing because they see the world through the same lenses, are subject to similar constraints and incentives, and thus feature stories or maintain silence together in tacit collective action and leader-follower behavior” (Herman and Chomsky 1988, p. lx). While no obvious advice is given on what to cut out, the journalists choose to practice self-censorship in a turn that counters the optimism of the Universal Declaration of Human Rights; we read the following on the 1959–1974 Vietnam war:

It is a highly significant fact that neither then, not before, was there any detectable questioning of the righteousness of the American cause in Vietnam, or of the necessity to proceed to full-scale ‘intervention.’ By that time, only questions of tactics and costs remained open, and further discussion in the mainstream media was largely limited to these narrow issues. While dissent and domestic controversy became a focus of media coverage from 1965, the actual views of dissidents and resisters were virtually excluded. These individuals were presented primarily as a threat to order, and while their tactics might be discussed, their views were not: ‘The antiwar movement stood at the bottom of the media’s hierarchy of legitimate political actors,’ Daniel Hallin concludes from his survey of television coverage (the print media were hardly different), ‘and its access to the news and influence over it were still more limited’. (Herman and Chomsky 1988, p. 172)

Selective treatment of the information by the mainstream media contributed to the creation of limits to the debate about the Vietnam war. Decision-making subsequently happened within such bounds; once legitimacy of the war effort stands as a nonissue, the problems of interest become those of tactical efficiency in the battlefield. By enforcing the limits on criticism consistently enough, the field of personal decision shrinks contrary to what would be expected if one reads the guarantees of freedom of expression. While the *First Amendment to the United States Constitution* tells that “congress shall make no law . . . abridging the freedom of speech, or of the press,” it is not the congress here, but the journalists who confronted the recommendations of freedom of expression to whatever other consideration they may have deemed relevant, and decided that freedoms should be curtailed. It is not unexpected that the subsequent view of the events is distorted:

From the point of view of the media, or ‘the culture,’ there is no such event in the US attack against the South Vietnam and the rest of Indochina. One would be hard put to find even an [*sic*] single reference within the mainstream to any such event, or any recognition



that history could possibly be viewed from this perspective — just as *Pravda*, presumably, records no such event as the [1978–1989] Soviet invasion of Afghanistan, only the defense of Afghanistan against ‘bandits’ supported by the CIA. . . There was much debate during the [Vietnam] war over whether the North Vietnamese were guilty of aggression in Vietnam. . . but there was *no discussion* of whether the United States was guilty of aggression in its direct attack against South Vietnam, then all of Indochina. These intriguing facts reflect the overwhelming dominance of the state propaganda system and its ability to set the terms of thought and discussion, even for those who believe themselves to be taking an ‘adversarial stance’. (Herman and Chomsky 1988, p. 184–185; Herman and Chomsky’s emphasis)

The three cases of censorship – 1930s Russia, France’s monarchy, and the Vietnam-era USA – illustrate three different relationships of individuals to meta-advice and the mechanisms of consent. In Stalin’s Russia, the censor follows advice or meta-advice on what is desirable and what should be suppressed. Broad parallels are visible between the practices of the church and king in France, and those of the Stalinist apparatus. Both seek consent through violent reprimand, on gallows and in the Bastille in France, the Gulag in Russia. Both dispensed the meta-advice of dictatorship, although at different times and with different outcomes. Both found many individuals willing to follow the advice to known effects. While an optimist might hold the Universal Declaration of Human Rights high in the face of authoritarian rule, possibly going as far as to claim that the contemporary world has learned from past mistakes, self-censorship in the mainstream media tells otherwise.

Despite the differences in kinds of meta-advice and advice that may be dispensed within various forms of government, the problem of how to assess advice remains at least equally important to the individual under a democracy as under a dictatorship. What is constant in both of these forms is their aim to coordinate, so that dispensing both meta-advice and advice is a necessity. Whether the purposes of that coordination are acceptable to the individual, and whether the intended coordination will take place is within the realm of personal decision, i.e., depends on the individual’s assessment of advice and meta-advice. That it is a personal decision does not mean that he is somehow a neutral analyst who rationally considers all his alternatives and chooses whether to follow or ignore advice only on grounds of its content. The examples above pointed out that advice should be distinguished from meta-advice in such analysis, but they also illustrated a distance between these two and choice. Advice alone is certainly not enough to predict choice, as the diligent Nazi dictator knew when he felt obliged to cancel the various freedoms through constitutional amendment in form, and violence in fact. We have observed that advice and choice do seem to interact, in the sense that behavior of decision-makers in laboratory games is different to a statistically significant degree when advice is present than when it is absent. Any assessment will require the knowledge of not only the advice and meta-advice that are dispensed, but also other characteristics of the advice-giving context. To be successful in guiding the individuals to act as advised, the Nazi dictatorship modified the legal and law enforcement framework, thus creating a particular system of incentives. By restricting the freedom of expression and prosecuting its practice, dictatorial rule clearly conveyed to the individual that the exercise of freedoms will be compensated by violence, while obedience



will give more desirable effects. In this way, it not only dispensed advice and meta-advice, but also aimed to influence its acceptance. We see then that any analysis of advice and meta-advice cannot focus solely on these two, but must be open to considerations that are aimed to regulate the ingestion of (meta-)advice.

## 2.6 Advice in Economic Coordination

Definitions of economics from the end of the nineteenth and the beginning of the twentieth centuries placed emphasis on the study of phenomena through prices, the explanation of the concept and causes of welfare, or of the cooperation methods people use to meet material needs. In *An essay on the nature & significance of economic science*, Lionel Robbins, a twentieth century English economist, offered a broad definition that is still often repeated today. “Economics” he suggested “is the science which studies human behavior as a relationship between ends and scarce means which have alternative uses” (Robbins 1942, p. 16). Some of the questions that we today see as clearly pertaining to economics were discussed as early as in Aristotle’s Greece, when he argued that coordination will be easier if the actions of the citizens are predictable. He saw shared virtues as necessary for predictability in affairs of exchange (Crespo 2008). While the complexities of managing trade and currencies were recognized in the period from the Gutenberg’s invention of the modern printing process in 1444 to the eighteenth century, economics develops as a proper field of inquiry of regularities with the physiocrats, David Hume and Adam Smith. In arguing that economic growth is determined by the land available for agriculture, the physiocrats recognized the importance of an appropriate, or optimal allocation of resources for growth (Muller 1978). There would be no worry about allocation if the resources were not rare. To justify private property, Hume tells us that it emanates from the scarcity of resources (Hume 1739–1740). Adam Smith talks of scarcity in relation to value: “. . .the value of . . .metals has, in all ages and nations, arisen chiefly from their scarcity, and that their scarcity has arisen from the very small quantities of them which nature has anywhere deposited in one place, from the hard and intractable substances with which she has almost everywhere surrounded those small quantities, and consequently from the labor and expense which are everywhere necessary in order to penetrate to and get at them” (Smith 1904, Book IV, Chap. VII).

When an individual calls onto the labors of others for the satisfaction of those needs beyond his own efforts, that is, when labor is divided and the economy specialized, some mechanism for the allocation of scarce resources is necessary. Not only will these scarce resources be the metals that Smith mentions above, but any result of labor, since most of the inputs necessary to labor are available in limited supply. Allocation and coordination go together only when the exchange of value takes place. In case a gift is given, that is, when the receiver of a scarce resource does not compensate the giver, there is no need for coordination other than what is necessary to arrange the transfer of the gift. When value goes both ways, so that

we can call the giver a seller and the receiver a buyer, allocation and coordination are inseparable. Since both participants in the exchange give away the products of their labors, each is forced to plan how to invest own resources. They are effectively asked to decide what, when, and in what quantity to offer and demand. Any such decision cannot be taken in isolation, independently of the information about the actions of the others with whom it may be relevant to exchange value. The mechanism of coordination will therefore dispense recommendations on courses of action. The allocation of resources will require the definition and distribution of advice to the individuals. We can consequently look at any mechanism for the allocation of scarce resources as a mechanism that produces and dispenses advice. We are in turn brought to ask what advice and meta-advice do dispense the usually discussed allocation mechanisms in economics, namely, allocation via central planning and via markets.

### ***2.6.1 Central Planning***

The growing material wealth of the consumers in the USA and Japan since the 1950s produced a rising demand for spiny lobster. While lobsters were not considered a delicacy in the nineteenth century, their culinary status kept improving (or worsening, as it depends on the perspective) and led countries rich in their habitats to organize catch and export. In a study of how the Cubans organized the fishing of the spiny lobster since 1975 (Joyce 1997), Ian Joyce, a geographer, illustrates the approach of central planning to the allocation of scarce resources and the associated task of coordination.

After announcing the intention to introduce economic efficiency into state enterprises, Fidel Castro's regime formed the Ministry of Fishing Industry (MFI). The Ministry took the responsibility of all planning, regulation, and production in all of Cuba's fisheries, which meant that it basically had to deal with all tasks that are related to the natural marine products. This included the direction of import and export, the actual fishing and distribution, and the buying and maintenance of all fishing equipment. From the moment the fish is caught, to the moment it is delivered to a kitchen or store, the beast goes only through the hands of the MFI employees. The most obviously scarce resources were the fishing zones and the fishing gear, including vessels. Since the MFI sought efficiency, it intended to avoid overfishing to protect the lobster stock. As a central planner, the Ministry split the fishing areas in zones and distributed the exclusive rights to these to the various local fishing enterprises. The exclusive rights made it unnecessary to improve the technology of fishing, for there is no competition; the clarity and shallowness of the water make the traps visible without advanced equipment on the vessels. There were in 1988 some 1,200 fishermen operating the vessels. Each received a base wage, owned neither the vessel nor equipment, and was eligible for bonuses if the goals of catch weight and quality were met; the goal depended on the previous 5 years' experience, gear, zone, vessel, species caught, and crew size. To regulate the fishing of lobsters, the

closed season was observed to reduce the interference with reproduction and allows growth. Inspections were also practiced to enforce the minimal legal size of the lobsters. The entire enterprise relied on the expertise of the Fisheries Research Center, which as a part of the MFI studied the fishing zones, the data from the fishing enterprises, and participated in the definition of regulations. A biologist worked at each lobster enterprise and liaised with the research center.

What we see from Joyce's study is that the MFI acts as a central planner in the principal activities involved in organized fishing. From the data on the technologies available for the task at hand, evaluations of the productivity of these technologies, the available natural resources and whatever other information its members deem relevant, it determines who will do what, how, and when in the catching, processing, and distribution of marine organisms in the Cuban territorial waters. In doing so, it dispenses advice to every fisherman it employs. Consent is obtained through the employment relation: the salary that the fisherman receives buys his consent, or equivalently, the absence of consent to the advice that the MFI dispenses means the absence of the salary.

If it is the advice of the central planner, combined with mechanism of consent that coordinate the activity, where is meta-advice in central planning? In discussing dictatorship, we extracted meta-advice from the ways it seeks consent to the advice the government dispenses. If the ways in which consent is sought are such that they ban the public assessment of the planner's advice, then the central planner effectively dispenses the same meta-advice as the dictator. Consider again the situation of the individual Cuban fisherman since 1975. We said that his consent to the planner's advice is sought through the salary. If we ask what would happen if he were to question the planner's advice, we can see clearer the planner's meta-advice. The salary is certainly not the only means for obtaining consent, although it is the most obvious in this case. Since the fishing zones are distributed and the MFI aimed at managing all activities related to fishing, the fisherman really has few options besides following the advice of the MFI. He could either engage in illegal fishing, or stop fishing altogether and work in another industry. Neither of the two options can be taken as public announcements of his independent assessment of the planner's advice. If he chooses to specialize in another industry, he will be in the same situation as long as that other industry is organized by a central planner. As the 1959 Cuban Revolution was a "communist revolution," we can safely assume that any industry other than fishing was centrally planned given that communism relies on such planning for coordination. A move between industries will not change the fisherman's ability to question the advice that a central planner gives, since he will simply replace one planner with another. If the fisherman chooses illegal fishing instead, he is not explicitly questioning advice. We can only conclude then that the central planner's advice cannot be questioned and thereby observe a parallel – in terms of meta-advice alone – of central planning as an allocation mechanism and dictatorship as a form of government.

In introducing meta-advice earlier, we said it is any recommendation that speaks about other advice. By observing that the central planner dispenses meta-advice that forbids the questioning of advice, we considered only one kind of meta-advice,

namely one that states whether the independent assessment of advice can be publicly performed and announced. We can of course be interested in other meta-advice.

Any way of allocating scarce resources must provide meta-advice about the distinction between the appropriate and inappropriate advice. Roughly speaking, there must be some meta-advice that says what qualifies as “good” and what as “bad” advice.

Consider the question of whether the Cuban fishermen should own the vessels and the tools of their trade. In the arrangement that was observed by Joyce, they do not. The fishermen only use the vessels and tools owned by the state fisheries, which begs the question of why the Ministry advises fishermen against private ownership. Equivalently, this is to ask what meta-advice the MFI follows to consider private ownership as inappropriate and recommend state ownership instead. Consider the following two hypothetical, but plausible explanations; it is actually unimportant whether they would effectively be given by the MFI if asked, as we are only interested in the differences between the two.

- Before the state enterprises owned and manufactured the vessels and tools, few fishermen could afford to buy, build, or rent them. The quantity of fish caught was significantly below what could be exported at profit and what was believed that the sea could provide without exhausting the natural resources. Consequently, the vessels and tools were produced by the state and remained in the ownership of the state, while the fishermen were employed to operate them.
- The Cuban revolution of the 1950s led to a form of government that intended to realize at least some of the ideas advocated by Karl Marx and Friedrich Engels in mid- and late nineteenth century. Che Guevara argued that the aim of the revolution was to fight against not only misery, but also “alienation” (Hollander 1988), whereby the latter term is used presumably in the sense that Marx gave to it. Before Marx, Rousseau, Hobbes, and Locke used alienation to describe a positive phenomenon through which the individual obtains personal freedom, while in twentieth century research in sociology after Marx, it has been associated with attitudes of loneliness, apathy, powerlessness, and loss of values (Twinning 1980). Although the term is not very consistently or clearly used in Marx’ manuscripts (Williamson and Cullingford 1997), alienation seems to designate both a state of an individual being removed from his work and the products of his work and the process by which this happens. The removal seems to occur because the (factory, assembly-line) worker is given such a limited responsibility in his repetitive task that he fails to express and develop own inventiveness and creativity and is distant from the final products of the labors to which he contributes. For Marx, alienation cannot be countered without the abolition of private property (Churchich 1990). His argument seems to be that if the tools of production are in private property, the worker does not decide how he will use these and will not own the product of his efforts. Distributing the tools of production to workers solves nothing in this perspective, since it allows any worker to engage another worker and place the latter in the same situation. As a consequence of this idea, private ownership of the fishing vessels and associated tools would lead to alienation, so that state ownership is recommended.

In the first explanation above, we could assume that some analysis was done of the fishermen's financial conditions and of the fishing resources to provide the basis for the argument: the fishermen cannot use to full capacity the natural resources (i.e., they could catch more if they had more vessels), and since they are poor, they cannot afford new vessels. Hence, the state chooses to manufacture and own the vessels. Observe that the meta-advice, which performs the discrimination between desirable (i.e., state-owned vessels) and undesirable (i.e., fishermen own the vessels) advice here, seems to be considerations of feasibility and efficiency. To recommend to fishermen that they should not own vessels and tools is explained by saying that they cannot buy or build them since they are poor (i.e., it is not feasible to have fishermen build their own vessels), and to leave the situation as-is would mean that the seas are under-exploited (i.e., it is inefficient to have a fishing industry, which under-exploits the resources). The meta-advice thus says that any solution that is feasible is preferred to any solution that is evaluated as infeasible, and any solution that is evaluated as more efficient than another is preferred to that other solution.

Meta-advice that recommends efficiency and feasibility over their opposites cannot be the only meta-advice that lead to the recommendation for state-ownership over another solution. A bank may estimate the future revenues from fishing as sufficient for a fisherman to pay a long-term loan for a vessel. A ship-builder would produce the vessels, the fisherman would purchase it with the money lent from the bank, and the bank would collect part of the revenues of the fisherman over some period of time. If the bank does not deem the fisherman worthy of credit, the state could intervene as a guarantor for the fisherman, but it need not own the vessels. We could invoke, say, the sustainability of the exploitation of marine resources: if the fishermen own the boats, then the state will have less control over how much is caught, when and by whom, being consequently obliged to police the seas.

We cannot identify in the first of the two explanations above the meta-advice that alone tips the scale in favor of state over private ownership of the fishing vessels. We hypothesized that efficiency, feasibility, and sustainability may favor the recommendation that state owns the vessels, given the situation of the fishermen. However, none of these meta-advice allow us to immediately favor ownership by the state. Things are much less open to interpretation in the second explanation, where the notion of alienation is linked to private property. There, meta-advice comes from Marx' manuscripts, as interpreted by the Cuban revolutionaries, and says quite clearly that private ownership should be avoided. Such meta-advice sharply discriminates between any advice that favors private ownership and some other arrangement: to accept such meta-advice leads the individual to reject any advice (be it about fisheries, factories, or agricultural land) that recommends private ownership of productive means. Observe the extent to which this sort of meta-advice simplifies decision-making: what matters only is that the individual doing the work is not the owner of the tools he uses, and that effectively no other individual owns them to the extent that he can employ other to work for him (since this will, if you accept Marx' argument, alienate the worker). What industry this may be, what natural resources it may be using, and what small or large investment the productive technology may require are all considerations that become entirely irrelevant

once the said meta-advice is adopted. Its adoption literally eliminates the necessity to think about any problem of private property at all. This does not mean that accepting meta-advice somehow solves the problem of whether the workers should own the productive means. Instead, that meta-advice places a number of alternatives (say, the private ownership of vessels, or of agricultural land) outside the bounds of decision-making: private ownership becomes an unacceptable alternative, whatever the industry.

Central planning in Cuba in the 1960s gives many more illustrations of meta-advice that aims to guide the evaluation of alternative advice. The Cuban economic system put in place in the 1960s through the Ministry of Industry, and referred to as the “centralized budgetary system of finance” had several important characteristics (Valdés 1979). The first is that centralization of decision-making was taken as far as it could go: after the means of production were socialized across all of agriculture and industry (i.e., taken from private and placed into state ownership without compensation to previous owners), all of economic planning, decision-making, and administrative tasks were placed under the responsibility of one authority. The second trait was the centralization of financial resources, so that every state enterprise withdrew the money it needed directly from the central bank and paid no interest on it; whenever the enterprise deposited amounts to its account, these automatically belonged to the state and could be withdrawn by any other state enterprise. Third, there were no cash transactions between enterprises. The exchange between state enterprises happens according to a plan and involves no cash. The entire economy roughly resembles a large firm, inside which no monetary transactions occur between departments. Fourth, prices were set by state planners by adding production costs (raw materials, wages, distribution, etc.) to a proportion of public investment in health, education, and so on. Finally, production quotas were set for all workers, and wages were paid in the proportion of the quota that the worker met. Che Guevara summarizes the aims of such policies as follows: “With respect to material interest, what we want to achieve with this system is to prevent the lever from becoming something that compels the individual, as an individual, or the collective of individuals, to struggle desperately with others so as to bring about certain conditions of production or distribution that would accord them special privileges” (cited in Valdés 1979, p. 15). As Nelson Valdés, a sociologist and historian of Latin America, observes in relation to this quote, the expressed aim is to build a society without relying on notions of market, profit, and material interest (Valdés 1979). Guevara’s meta-advice is that any advice that fosters competition between individuals is unacceptable. The Cuban economy did not remain such for a long period, as the Che Guevara’s centralized budgetary system of finance was replaced by Fidel Castro’s decentralized budgetary system, in which there was more emphasis on local state enterprises, in that central planning was replaced by regional planning, and prices were no longer set on the basis of production costs and similar considerations; they were revised to make the bare necessities more accessible to the population, and some products were distributed free of charge. Wages were no longer paid according to the quantity or quality of output, but were left for worker assemblies to decide. After 1968, education, health care, housing, and childcare were all free, and there

were no taxes. Matters changed subsequently again, as we saw from the fishing enterprises, who did pay wages according to quotas and did provide incentives as bonuses. Advice and meta-advice changed in each period.

### 2.6.2 *Market*

At the annual meeting of the Boards of Governors of the World Bank Group and International Monetary Fund on 29 September 1981, Ronald Reagan, in the role of the 40th president of the USA at the beginning of his first term in office spoke of the market as of a mystifying phenomenon: “The societies which have achieved the most spectacular broad-based economic progress in the shortest period of time are not the most tightly controlled, not necessarily the biggest in size, or the wealthiest in natural resources. No, what unites them all is their willingness to believe in the magic of the marketplace.” Although this last remark is a figure of speech, it is actually not out of place. The term “market” is typically used to refer to a place where buyers and sellers can exchange goods and services. This mechanism of coordination has developed over (at least) the last seven millennia, from the physical trading places where different tribes and members of early civilizations met, to the modern markets that do not require the physical meeting of the trading partners. The relevance of a physical location for a market seems to vanish together with the hopes of a theory of markets that would incorporate various insights offered mainly in economics, sociology, law, and history into the functioning of this phenomenon (Swedberg 1994). It is hardly only Reagan who was puzzled by the workings of markets.

When we contrast a market with a central planner, such as that of the Cuban economy, what appears puzzling is that a market manages to coordinate buyers and sellers without an identifiable central planner. Coordination through the market is not based on the choices of a single decision-maker, but occurs through the interplay of decentralized decision-making of potentially many buyers and sellers. While it is clear who dispenses the information relevant for coordination in the case of the central decision-maker, who dispenses advice seems less obvious in the market.

For the economist, a key feature of the market is that it sets the price at which the buyers are willing to buy and the sellers are willing to sell. At such a price, the market is said to clear. In the essay on *The Nature of the Firm*, Ronald Coase (1937), a British economist cites Hayek to illustrate a way to think of the economy as of a system coordinated by prices, and explains:

... in economic theory, we find that the allocation of factors of production between different uses is determined by the price mechanism. The price of factor A becomes higher in X than in Y. As a result, A moves from Y to X until the difference between the prices of X and Y, except in so far as it compensates for other differential advantages, disappears. (Coase 1937, p. 387)



What Coase describes is the process that leads a profit-seeking individual to reallocate factors of production whenever the opportunity for higher profit presents itself. If he can make more out of A in X than in Y, the economist expects him to move from Y to X. The individual, an entrepreneur seeks in prices the coordinating information, which he subsequently uses in choosing how to allocate own resources: “In drafting their plans the entrepreneurs never make these prices enter into their calculations without paying regard to anticipated changes. The prices of the immediate past are for them only the starting point of deliberations leading to forecasts of future prices” (von Mises 1966, p. 336).

To the extent that a price recommends a course of action, it can be thought of as advice. If we take, as Coase above, that the individual aims for profit-maximization, then the current price of, say an input to his production process will be a relevant starting point to estimate his production costs. The current price will be necessary to decide whether to buy, and this whenever the resources available for the purchase do not render the price completely irrelevant. Of course, we need not go so far to see what kind of action a price recommends. If we make the less ambitious assumption that the individual simply thought before coming to the market about the maximal price he is willing to pay, then the market price will advise to buy if the market price is below the maximal price the buyer is willing to pay; otherwise, the market price advises the buyer against the purchase. It is not necessary to assume profit-seeking for the market price to act as advice.

While the meeting of the buyers and sellers, and the setting of a market-clearing price may be enough to depict the basics of a market, much more is needed for the actual exchange to take place. Early development of commercial law is illustrative in this respect (Swedberg 2003). With the rapid growth of the Western agriculture and trade in the eleventh and twelfth centuries, merchants organized markets and fairs and in the process developed their own laws to regulate the buying, selling, transport, and insurance of goods. As Harold Berman, an American legal scholar, argues, a central element of the mercantile law was the principle of reciprocity. The buyer parts from money, and the seller from the goods sold, while both expect to be better off after doing so. Moreover, there must be some equality in burdens and benefits, which as Berman observes has two additional aspects. The first requires that both parties enter the exchange “without duress or fraud or other abuse of the will or knowledge of either party” (Berman 1983, p. 344). The second requires that the exchange must, even if entered with will and knowledge, neither impose on a party the costs that are significantly disproportionate to the expected benefits, nor be disadvantageous to third parties or society. Rules were enforced via courts formed at markets and fairs. Judges were merchants elected by their fellows present at the market or fair. While price may be one advice dispensed by the market, the legal codes surrounding the market enable its existence by advising the parties on how to enter and relate to each other in a transaction. Enforcement then aims to ensure the consent to the legislated advice.

To see that prices provide one kind of advice on the market and that the market’s legal cadre — say, *lex mercatoria* from eleventh century on — and the prices as



advice begs the questions of if, how and what meta-advice may be dispensed on a market along with prices and legal recommendations.

Meta-advice dispensed with the recommendations of *lex mercatoria* is visible from our earlier discussion. Since the early commercial law places the principle of reciprocity in exchange as its central element, it must dispense meta-advice that will, given advice that differ in terms of the fit with the chosen notion of reciprocity, recommend one of these alternatives. If the seller sets a price that is significantly disproportionate to the benefits that the buyer should expect, the meta-advice that arises out of the principle of reciprocity will state that this price is an inappropriate advice since it violates reciprocity.

Consider then the meta-advice that may speak of prices. Any exchange involves a bidirectional transfer of value, regardless of what may be exchanged. If not, then either there is no need for exchange in the first place, or the basics of *lex mercatoria* would be violated (such as when value moves only in one way). If we then take the mainstream perspective of contemporary economics, which, as Coase observed (Coase 1988), is predominantly concerned with the role of the market as a way to set prices, then the value that any party in the exchange gives away is measured by the price. This does not mean that the value perceived by some individual equates to the value that the price estimates, but only that the market effectively provides no other measurement of value. Where is meta-advice then? Since the market sets a price at which the exchange happens, the market states two recommendations about the price itself: (1) instead of any other price, use the market price for the transaction, and (2) in place of any other price, use the market price as an estimate of the value of the product or service being exchanged. It is trivial to observe that no market will coordinate the individuals if these two meta-advice are not accepted by these same individuals. Rejection of the first meta-advice may lead the market participants to fix the price in a different way than the market, say by private negotiations, or by way of auction. If some individuals reject the second meta-advice, they will need to seek another way to estimate the value of the products and services that they are buying and selling. They may consider that the price should be determined to render the product or service accessible to most, as Fidel Castro did when his economic policies reduced the price of bare necessities regardless of their production costs, or tie the price strongly to the production costs, as Che Guevara did in the Cuban economy before Castro took over.

To reject the price that is set by the market does not necessarily mean that we reject the meta-advice dispensed by the market. A price may be contested on grounds that the market is distorted, as when there is one monopolist seller or few sellers who form an oligopoly, that is, set the market price by an agreement between them. Meta-advice remains standing in such cases, since the aim of the criticism of price may be precisely to eliminate the actual advice that contradicts the accepted meta-advice: a price that is set by agreement of several sellers is effectively not a market price formed by the choices of many different sellers and buyers.

## 2.7 Homo Follis

As one of the more interesting movies of the 1990s, many remember rather fondly the 1994 comedy drama *Forrest Gump* for Tom Hanks' depiction of its main character of the same name, a naïve anti-cynic, hopeful against all odds, whose paltry IQ is no obstacle to a multicareer lifetime of sports stardom, warfare heroism, shrimp tycoonism, and fatherhood. We see through him a counterintuitive, though soothing image of the unlikely American Dream, in which a pragmatic and silly everyman meets wealth and shows arresting insight as an accidental pop-philosopher, who manages to convey folkloric wisdom in such a legendary phrase as “sh—t happens” after stepping into excrement on a transcontinental run without a cause. One of his less-appreciated talents seems to be the exceptional absence of a critical attitude toward the recommendations he is given. At some point in the story, the movie script puts him in the army, facing a shouting drill sergeant, the latter resembling here a politically correct and less expletive-laden tribute to Stanley Kubrick's *Gunnery Sgt. Hartmann* of *Full Metal Jacket*, a 1987 war drama:

*Drill Sergeant:* Gump! What's your sole purpose in this Army?

*Forrest:* To do whatever you tell me, Drill Sergeant!

*Drill Sergeant:* Goddamnit, Gump! You're a goddamned genius! That's the most outstanding answer I've ever heard. You must have a goddamned I.Q. of a hundred and sixty! You are goddamned gifted, Private Gump!

*The Drill Sergeant moves down the line to the next man.*

*Drill Sergeant:* Listen up, people. . .

*Forrest (voice over):* Now, for some reason, I fit in the Army like one of them round pegs. It's not really hard. You just make your bed real neat and remember to stand up straight.

*Drill Sergeant:* That is one very intelligent individual! You lock your scuzzy bodies up behind that private and do exactly what he does and you will go far in this man's army!

*Forrest (voice over):* And always answer every question with “Yes, Drill Sergeant!”

*Drill Sergeant:* Is that clear?

*Forrest and recruits:* Yes, Drill Sergeant!

The scene is one of many that highlight a recurring paradox in the movie, where the unintelligent character unexpectedly goes straight to the point, a surprisingly wise everyman of sorts. Forrest seems to have got the spirit of the military service right from his very first days in the barracks: he acknowledges that he will take any advice from his superior, and more importantly, he does so by acknowledging his acceptance of the meta-advice that the military dispenses. Do as you are told and confirm that you understand what you are told so that others know what you will do. Mechanistic coordination at its best. The scene is of course ironic, as it should be. Nevertheless, it helps us here to point out again the distinction that exists between advice and meta-advice, and the basic argument that this chapter substantiated: namely, that how an individual assesses advice is a determinant of his relationship to and position within societies' coordination mechanisms and thereby within the political, economic, legal, and other realms.

That Forrest accepts the military's meta-advice does not mean that he cannot reflect on the merits of such a recommendation even as he follows it. The individual who does not exercise an independent assessment of advice and meta-advice that he receives is not so different from a Forrest Gump without a movie script. Forrest can passively take events as they come, for it is the script that organizes the consequences of his choices, guaranteeing him success after occasional failures so that the audience takes away some pleasure out of the viewing. We know even as we watch the movie for the first time that the army will not take him to injury or death, but we know nothing of the sort for any real-world recruit entering the compulsory military service in, say Israel or Colombia. In the absence of the script, a Forrest Gump is reduced to what the cynic sees all along in the main character of the movie, namely naivety and intellectual passivity. Neither is laudable.

It is hard to say if Carl Linnaeus was being ironic when he coined the term *Homo sapiens* in the mid-eighteenth century, the Latin for wise/knowing human being. It is a name appropriate neither for a scriptless Gump nor the destructive revolutionary living the illusion that a few cuts of the guillotine, or a red booklet will magically alleviate all that she may deem undesirable in either thoughts or acts of others. If neither of these supposed extremes warrants Linnaeus' designation, is its object to be sought somewhere along the continuum? The usual response would be to say that things are complex, that any account of any position along that fictive line would only be a crude generalization. But what if there is really no complexity, that the two are no opposites at all? Undoubtedly different in many respects, the Gump and the revolutionary are nevertheless very close in that one crucial respect: both will do just as they are advised – where Gump follows the Drill Sergeant, the revolutionary obeys the spiritual or intellectual guide. If the latter complains to this trivial observation, he should consider another trivial observation – hardly one lacking material evidence – that many a tyrant started off as a dedicated and seemingly benevolent revolutionary, a critic of the ideas and deeds preceding his own.

If not cases of *Homo sapiens*, what are the Gumps and the seemingly intellectually and otherwise sophisticated – pick your own, one or more of – Capitalist, Liberal, Ecologist, Marxist, Maoist, Libertarianist and so on, if not instances of *Homo follis*?<sup>5</sup> That of course is a generalization, but one that does not miss the point: the interesting opposition is not between the Gumps and the revolutionaries, it is between both of them – *Homo follis* – and perhaps *Homo diligens*, from Latin for careful, diligent. Careful and diligent in the assessment of advice he dispenses and accepts, a *Homo diligens* would be an individual who does such an assessment not from the position of the absolutely independent and inactive critic, but while understanding that coordination, and thus the taking of advice is required for

---

<sup>5</sup> “The pejorative nature of the term fool is strengthened by a knowledge of its etymology. Its source, the Latin word *follis*, meant ‘a bag or sack, a large inflated ball, a pair of bellows.’ Users of the word in Late Latin, however, saw a resemblance between the bellows or the inflated ball and a person who was what we would call ‘a windbag’ or ‘an airhead.’ The word, which passed into English by way of French, is first recorded in English in a work written around the beginning of the thirteenth century with the sense a foolish, stupid, or ignorant person.” (Anonymous 2000a)

coordination to happen, and some of his own desires to be satisfied. As no advice is in itself closed to critique, and any advice and coordination mechanism is amenable to change, the *Homo diligens* position remains a tenable one, thereby justifying the pejorative sense of *Homo follicis*.

That the *Homo diligens* may be a desirable position is not so common a thought as it may seem. The dictatorship, the polyarchy, and whatever is in-between all coordinate more easily when dealing with the *Homo follicis*. To systematically kill, as the Nazi party did in the Second World War, either Gumps or revolutionaries will do, but the latter will be more efficient: the former will simply obey, while the latter will work with zeal, believing in the righteousness of the act they have been advised to perform. But this is hardly confined to reigns of terror (e.g., Herman and Chomsky 1988): as John Locke, a seventeenth century English philosopher, wrote, “you may as soon hope to leave all the day laborers and tradesmen, the spinsters and dairy-maids, perfect mathematicians, as to have them perfect in ethics this way: having plain commands is the sure and only course to bring them to obedience and practice: the greatest part cannot know, and therefore they must believe” (Locke 1696). Interestingly enough, he was a philosopher of Enlightenment, one of the promoters of the notion that a government cannot rule without the consent of the governed. While there must be consent, it seems, that consent will need to come out not of diligent reflection, but from injunction. Facing an injunction, *Homo follicis* will blindly follow, while a *Homo diligens* would reflect to determine the extent to which the advisor can be influenced, his advice changed, all the while the advised will remain well aware of his interest in coordination to take place toward the achievement of his own aims.

The division of physical and intellectual labor continues to advance, and specialization follows. An individual who specializes acquires knowledge that is most relevant for the decision situations related to his specialty. As a consequence, the specialist will revert to the generalist, a layperson any time he is asked to choose outside of his own specialty. Any variant of polyarchy will ask the specialist to revert to the generalist at least any time there are elections, or some tool of direct democracy is used. To the specialist in coordination, say, a politician, a manager, the problem then seems to be one of knowing whom he is governing, *Homo follicis* or *Homo diligens*, and if one follows Walter Lipmann’s arguments, to transform the latter into the former:

In the absence of institutions and education by which the environment is so successfully reported that the realities of public life stand out sharply against self-centered opinion, the common interests very largely elude public opinion entirely, and can be managed only by a specialized class whose personal interests reach beyond the locality. This class is irresponsible, for it acts upon information that is not common property, in situations that the public at large does not conceive, and it can be held to account only on the accomplished fact. (Lippmann 1922, p. 195)

To the other specialist, the one that the politician sees as the subject, the manager as the managed, the problem is very much the opposite and a difficult one. Her reversal from the specialist to the generalist will take place much more often, as many choices other than those related to the individual’s specialty are continually

expected. In any such situation, advice and meta-advice will be available and they need to be assessed.

To recognize the unstoppable advancement of specialization is to see that the pressing issue becomes how an individual can analyze advice and meta-advice when these are beyond his own specialty. Stakes are not negligible, at play is his status as either a *Homo follis*, a *Homo diligens*, or perhaps even a *Homo sapiens*. Active participation in the game requires an understanding of the interplay between the advisor and the advised, of the difficult and hardly obvious issues that arise in that interaction, and of the fundamental tradeoff between the necessity to coordinate and really the only way to seriously approach advice giving and taking, that is, by a thorough, careful, diligent, and deep analysis of both the advice one dispenses and receives. The aim is to spell out the foundations of such an analysis in the ensuing chapters.

## Chapter 3

# Conceptual Analysis of Advice

Most people only very rarely know exactly what they are talking about. If this were so, then it would only ensue that their advice is very rarely worth much.

To claim this may seem too far off the mark and is certainly a very pessimistic stance. It is also a frightening thought, for it is soothing to believe that, if one knows little and thus perhaps controls little, someone else certainly must know and have matters under control: when the specialist reverts to the generalist, he will certainly hope – and act on that hope by seeking advice – that there are specialists for decision settings other than those of his own specialty. The inclination to believe so comes out of the obvious need for predictability, especially when coordination is concerned and aimed at, through the giving and taking of advice.

It is evidently very hard to uphold the claim that very few and only rarely know exactly what they are talking about. If anything, current state of science and technology offers striking counterarguments, given the extent that many natural and social phenomena obtain plausible explanations, some are predictable, and even amenable to influence and control. At least some must therefore know precisely what they are doing and advising, and it is presumably their deep specialization that shines when this is the case. Yet specialists they are, and they consequently still need advice in any decision situation outside their own specialty, whenever they revert from the role of a specialist to the role of a generalist.

Beyond the awareness of one's command of a problematic, the trait of an honest specialist ought to be the awareness of the limits of his expertise in that specialty. It is these limits that produce the nuance in the advice that the specialist dispenses, when the topic of that advice effectively is his own specialty. And therein lies the trick: if the specialist nuances the advice that he gives precisely because he is aware of his own ignorance, then should he really succumb to the soothing conviction that others do the same when advising him? When he reverts to the specialist, and demands advice, is the advice he receives also as nuanced and carefully designed as that which he provides, or at least believes to provide?

To answer the positive is to ignore a simple fact about advice: one need not be educated at all in the analysis and the design of advice to produce and dispense advice. Requiring only at least two people who can communicate, advice-giving is certainly a very accessible activity, one that often requires very little material and intellectual investment. Given this ease with which advice can be produced and

dispensed, it does not seem too far off to say that advice will very often be irrelevant, incorrect, imprecise. Again then, to answer the positive is to engage in some wishful thinking, namely, that most people most of the time do make the exceptional material and intellectual investment that is effectively necessary to conscientiously and meticulously craft the advice they dispense.

Even when naïveté remains an option, to expect only the best and brightest advice still cannot avoid some serious practical problems. We can roughly call these the problems of communication. There do exist cases in which these problems are absent: e.g., suppose that the advisor is standing next to the gambler who is about to make a bet, and that the advisor knows in advance which option will win; it is enough to the advisor to point to the winning option for the recipient of advice to know what he has been recommended to choose. What is typically referred to as the meaning this advice is unequivocal, it admits no doubt or misunderstanding. If the naïve gambler here takes this advice, he will do well. That setting stands in stark contrast to a different, yet not uncommon situation: when we are advised by the media to believe that some recently invaded country at the outset of the twenty-first century has become a democracy, what is it precisely that we are advised to believe? If an enthusiast says that this is precisely what we should believe, namely that “that country now is a democracy”, then we may rightfully ask what could this possibly mean. A reply could be that, well, compared to conditions before an invasion, it is now certainly (more?) democratic. This sort of question and answer could go on *ad nauseam*. The point is that no amount of goodwill one ascribes to the advisor will resolve some classical and recurrent, and it seems unsolvable problems of communication. In the case of a paradoxically named democratic, yet occupied country, the question is what the meaning the term *democracy* has – or in other words and as we will see in this chapter, what the term *refers to*.

That it is desirable to know precisely which concepts and/or phenomena a term, or more broadly a *sign* is referring to, does not entail that this is the case for the many signs in general, as well as those used in giving advice. Broadly speaking, “meaning” may depend on other signs being used, on the situation in which they are used together, the people involved in the communication, these people’s beliefs, desires, intentions, emotions, moods, feelings – “meaning” is context-dependent and often to a considerable extent something that remains variable and unclear. It is often *all but* sharp and clear. To claim that it is clear and precise what one should believe when advised to believe that some country is a democracy, is to ignore and take as irrelevant the debates on what the term democracy refers to, what properties a political system must have to be referred to by the term *democracy*. It is to ignore a debate that is ongoing, and that has been revisited repeatedly at least since Pericle’s Athens. It is, in other words, to accept an exceptionally simplistic and uninterested perspective, hardly a laudable view.

What then remains of the claim that most people rarely know exactly what they are talking about? It certainly seems too general to be of much relevance. It is a sweeping claim that can straightforwardly be countered via other generalizations, say, that technological and scientific progress indicate precisely the opposite. Yet to accept its opposite in the face of such general counterarguments would still be

premature. Is it really so obvious, as the opposite would claim, that most people often know exactly what they are talking about? The truly important question is neither of these two, specifically because their answers are irrelevant to the individual facing a decision situation in which he has not the role of the specialist, but of the generalist who receives advice on how to act. In such a setting the problem really is not whether some people, some of the time know what they are talking about.

The arresting problem is instead this: *how does the specialist-turned-generalist, in the particular decision situation he has found himself in, establish if the advice he receives is relevant, sharp, clear, precise, or the opposite of those?* To answer this by taking seriously any of the two generalizations – that either most people know most of the time what they are advising or its opposite – is no different from tossing a coin: hardly any thinking is involved. Even when the advisor is a benevolent expert intent on doing his best when designing the advice he has, the understanding of the advice and the potential actions that the recipient of advice may take depend very much on the recipient: even when the advisor does know precisely what he is talking about, the recipient still may misinterpret what the advisor is talking about.

How a specialist-turned-generalist may establish whether the advice he is given is relevant, sharp, and so on is a question that bundles together two problems that should be treated separately: (1) *what is the advice referring to within the context of the specific decision situation?* and (2) *what is the purpose of that advice in that context?*

Both the individual who designs advice to dispense it, and the one on the receiving end who analyses advice, need to consider these questions. They will do so in different order and ways, however. For the advisor who does the design, the problem is to choose what signs to use in advice (i.e., what terms, phrases, and so on) to ensure that the aim he has by giving that advice can effectively be reached: for the advisor, the first problem is to decide on the aim of advice he will give, and then on what content and form to give to that advice. In contrast, the first problem for the recipient of advice is to establish what the advice he has received refers to in his specific decision situation, and then reflect on whether to accept or reject it; this second problem is one of finding the purpose of that advice and seeing if that purpose fits those of the recipient.

This chapter is concerned with how the first question may get answered, namely how it can be established what the advice is referring to within the context of the specific decision situation.<sup>1</sup> To put it bluntly, this chapter focuses on why it is difficult to make sure that the recipient of advice knows exactly what the advice is about. It is the premise of the conceptual analysis of advice that producing advice that is clear, precise, and relevant requires an understanding of such difficulties, of the mechanisms that produce these difficulties, and of the potential solutions thereto. Such an understanding requires that we look into how signs used when advice is communicated are defined by the advisor, and how the advisor can influence what the

---

<sup>1</sup> The second question which is to find out the purpose of advice in a given context is the topic of the fourth and later chapters.



recipient understands from these signs. This in turn makes it necessary that we discuss what these signs and the objects and concepts to which the signs refer may be, and how those reference relations between signs, concepts, and objects get formed. Conceptual analysis of advice looks into these concerns, assembling together insight and perspective primarily from philosophy and another interdisciplinary field, conceptual modeling, which has been concerned among others, with the modeling of information subsequently recorded and manipulated by computers.

Conceptual analysis of advice is very far from a cooking recipe, a definite sequence of well-defined tasks that may leave little to the chef to decide. It is rather a collection of guidelines and most prominently questions and problems, the answers to which have hardly been settled once and for all. Questions are general, they can be asked for any piece of advice one designs or receives. Guidelines are general as well, and their following can also be considered for any advice.

This chapter outlines how conceptual analysis proceeds for some given sign, e.g., a word. The chapter reflects what one needs to know to perform a conceptual analysis, and then illustrates how to mobilize this knowledge to perform a conceptual analysis of the word *advice*, the result being a definition of the term *advice*. In other words, the result of the conceptual analysis performed in the chapter is advice on what the term *advice* “means.” The result is, so to speak a piece of advice on what *advice* is. The chapter thus illustrates the steps that someone designing advice on how to analyze advice can take to move from a term that seems simple enough to its definition, and consider the problems that need to be resolved along the way, to provide a sufficiently relevant, clear, and precise definition.

To convey the meaning we intended for the term *advice* in the second chapter, we called upon the *Oxford English Dictionary* and *WordNet*. In doing so, we introduced a lexical shortcut: *advice* was equated with a phrase or sentence, whereby the aim of the latter two was to place constraints on the reading of any subsequent mention of the term *advice*. Whenever *advice* was written, it was assumed that the word is read according to the commonsense definition given in one of these sources. Going about a discussion in this way corresponds to the usual usage of the word *definition*; we “defined” advice by equating the word with a phrase or sentence, whereby the role of the latter is to convey the meaning that is subsumed under the term we were brought to define. The *Oxford English Dictionary* thus equates the word *advice* with several words, phrases, and sentences we mentioned earlier (see Sect. 2), while *WordNet* instead equates it with the phrase *a proposal for an appropriate course of action*. A reading of the definition from *WordNet* tells us that anything called *proposal* about a *course of action* that is *appropriate* is *advice*. We may stop there only if we are content with the meaning we associate with all words and phrases that are italicized in the past sentence. Otherwise, one solution is to look up the quoted words and phrases in a dictionary, then look up the words that appear in their definitions, and so recursively on. This is not unlike what an individual does when learning the vocabulary of a foreign language using a thesaurus in that language; start with a few words relative to the body of words in the foreign language, then seek to give meaning to new words by relating them to those few from the outset.

As John Locke observed in *An Essay Concerning Human Understanding*, working our way through definitions by recursively looking to define terms that appear therein will lead us to ask “where at last should we stop?” (Locke 1825, p. 307). In practice, the stopping criterion must be argued for within the context of interest, as an universal such criterion has been elusive. A commitment is thus typically made on some primitive terms, which then remain undefined, while their remaining undefined is argued for. The definition is built by relating these primitive terms. These are the rough outlines of the approach we shall take below to provide a definition of *advice*. It is important to emphasize that it is *a* definition of advice that is offered, not *the* definition thereof. *A* stands in place of *the* because the definition we will arrive to is relevant in the context of this text, and thereby not some sort of universal, context-independent definition.

The very intention of providing a definition of advice requires an engagement in the tricky business of defining. The present chapter has consequently two aims. The first is to reflect on the problems of designing definitions, and thereby on the desirable and undesirable characteristics of definitions. This is crucial not only because of the second aim, which is to offer and discuss a definition of advice, but because any advice will mention terms, themselves subsuming definitions. An analysis of definitions is thus one way to analyze advice, by breaking it down onto individual terms and phrases, then looking into the definitions thereof to elucidate the purposes of, and assumptions behind a piece of (meta-)advice. This two-step approach, first on definitions, then on a definition of advice has two important traits. The first is that it facilitates the criticism of the proposal for the understanding of *advice* that will be given below. It is effectively difficult to criticize and revise a proposal for anything, including this very proposal for the analysis of advice, if a considerable part of it remains hidden, that is, if its hypotheses are not explicitly stated to the extent allowed by the various resources that are available. We cannot seriously claim to offer an analysis of advice, if we cannot provide an analysis of the basic terms, on which the assessment of advice builds upon. The second important trait of this approach is that an explicit account of conceptual commitments illustrates in itself the difficulty there is to come to an agreement about the meaning of some primitive terms put together toward a definition of advice. In doing so, we rejoin the idea mentioned above: that a discussion of the design of definitions illustrate one kind of analysis, conceptual analysis that can be applied whenever we are interested in defining a term mentioned in some piece of advice or meta-advice.

## 3.1 Rigorous Definition

### 3.1.1 *Ostensive and Intensional Definitions*

It is not uncommon to find reflections on what it may be to define, and what a definition may be whenever there is an attempt at rigor in a discussion. In *Confessions*, Augustine’s autobiographical text written toward the very end of the fourth

century, he recollects how he went about learning words and their meaning at a very young age:

When they [his elders] called some thing by name and pointed it out while they spoke, I saw it and realized that the thing they wished to indicate was called by the name they then uttered. And what they meant was made plain by the gestures of their bodies, by a kind of natural language, common to all nations, which expresses itself through changes of countenance, glances of the eye, gestures and intonations which indicate a disposition and attitude — either to seek or to possess, to reject or to avoid. So it was that by frequently hearing words, in different phrases, I gradually identified the objects which the words stood for and, having formed my mouth to repeat these signs, I was thereby able to express my will. Thus I exchanged with those about me the verbal signs by which we express our wishes and advanced deeper into the stormy fellowship of human life, depending all the while upon the authority of my parents and the behest of my elders. (Augustine 1963, Book I, Chap. VIII)

In the terminology of studies concerned with the act of defining and of definitions, what he calls “name” is called *definiendum*. It is what those who read from left to right usually find on the left-hand side in the lists of definitions of a dictionary, that is, the word or phrase being defined. If we look at the method of defining that he describes, we see that it is through examples that he learns to associate names with things, that is, to relate a *definiendum* to a *definiens*. He is shown an apple and is told the word “apple”; this act leads him to associate the word apple with the object shown. This is what Ludwig Wittgenstein, a twentieth century philosopher, calls the ostensive teaching of words, which aims to establish a connection between a word and a thing (Wittgenstein 1953, paragraph 6). To perform ostensive definition is to relate a *definiendum* to a *definiens* by pointing out examples, or as Wittgenstein suggests: “the ostensive definition explains the use — the meaning — of the word when the overall role of the word in the language is clear. Thus if I know that someone means to explain a color-word to me the ostensive definition ‘That is called ‘sepia’ will help me to understand the word’” (Wittgenstein 1953, paragraph 30).

Ostensive definitions, just as any other sort of definition, come together with a number of significant questions, which Wittgenstein conveniently summarizes in saying that “you can say this [that the color you pointed to is sepia], so long as you do not forget that all sorts of problems attach to the words ‘to know’ or ‘to be clear’. One has already to know (or be able to do) something in order to be capable of asking a thing’s name. But what does one have to know?” (Wittgenstein 1953, paragraph 30).<sup>2</sup> An enthusiast could hope to define the word advice in an ostensive manner, by pointing to specific advice and saying that they are to be called *advice*. While we had started to some extent to do so in the second chapter, when we looked at the advice and meta-advice dispensed toward political and economic

---

<sup>2</sup> We shall discuss neither this question he asks, nor its potential answers. We shall leave this issue open, and assume for now that what one has to know is whatever is necessary for meaningful communication to occur. The communication will be meaningful only to the extent to which we can — imperfectly of course — exchange hints about the meanings intended for the words we use. We will come back to this in discussing the *context of reference*, later in this chapter.

coordination, aiming for an ostensive definition of advice is rather impractical: How would you know if something that has not been listed in that definition is not advice?

We may instead be interested in designing an *intensional definition*. Such a definition gives the necessary and hopefully sufficient conditions for a thing to be called by a particular name. Definiens thereby gives the characteristics that a thing *must* have to be called with the definiendum. An ostensive definition of natural numbers will thus always be incomplete: if we point to the signs 0, 1, and 2 written on a piece of paper, and say that these are natural numbers, what should the person we are teaching say when confronted to the sign 3? Instead (or together with) an ostensive definition, an intensional definition of natural numbers, then becomes relevant. One among other such definitions is that anything called a *natural number* is a member of the set possessing all the following properties (Landau 1951, pp. 1–18):

1. 1 (read: one) is a natural number.
2. For each natural number  $x$  there exists exactly one natural number, called the successor of  $x$ .
3. There is no natural number whose successor is 1.
4. For any given natural number, there is either no natural number or exactly one natural number whose successor is the given number.
5. Let there be a set  $M$  of natural numbers, with the following properties:
  - (a) 1 belongs to  $M$ .
  - (b) If a natural number belongs to the set  $M$  then so does its successor.

The above are called the Peano axioms, after the nineteenth to twentieth century Italian mathematician Giuseppe Peano. To say that a natural number is only that which satisfies the above properties is to say that it satisfies the above axioms. For the definition of natural numbers via Peano axioms to be meaningful, the individual will need to have already encountered the following: (1) an ostensive or intensional definition of *set*, (2) the signs for the ten Arabic numerals and how they are usually combined (e.g., loosely speaking, writing the sign for 1 just before the sign for 2 gives a sign that reads 12), (3) a definition of equality (otherwise, the fourth axiom will not make much sense), (4) a standard definition of the *total order* in set theory,<sup>3</sup> and (5) an understanding of what recursion is in mathematics (which is needed to grasp the fifth axiom). Few children will be taught what *natural number* means in this way. It is likely that ostensive definitions will be used, with exercises in the writing of number signs, and so on, without going into the formal mathematical (and rigorous) definitions of total order, set, and whatever else must be known before one attempts to grasp Peano axioms.

---

<sup>3</sup> The notion of successor has no clear meaning without a total order on the set  $M$  in the given axioms. Recall that some set  $S$  is totally ordered under the binary relation  $\leq$  if and only if any three members  $a$ ,  $b$ , and  $c$  of  $S$  satisfy all the following properties: (1) if  $a \leq b$  and  $b \leq a$  then  $a = b$ ; (2) if  $a \leq b$  and  $b \leq c$  then  $a \leq c$ ; and (3)  $a \leq b$  or  $b \leq a$ .

### 3.1.2 *Primitive Terms*

Apart from illustrating the distinction between defining something ostensively or intensionally, the definition of natural numbers via Peano axioms highlights the difficulties that may be encountered when definitions are designed. It is safe to say that any definition will make assumptions, which itself cannot make explicit, if only for the economy of exposition. Such is the case of the Peano axioms, which build on knowledge that must have been previously acquired. Recall the definition of a polyarchy (see Sect. 2.5). The conditions given for a polyarchy are not unlike the axioms above, as both define conditions that must be satisfied for the definiendum to apply. If some form of governance satisfies all conditions in the definition of a polyarchy, we can call that form of governance a polyarchy. Just as Peano axioms assume some prior definitions, so does the definition of polyarchy. We find therein many terms, such as constitution, governmental decisions, elected official, office, frequent, fair, free, and so on. Some understanding of all these is needed to grasp what the term polyarchy may mean. It is further clear that if two people understand any one or more of these terms differently, they will understand the term polyarchy differently.

To seek agreement on a definiendum requires the recursive definition of the terms appearing in its definiens. In addition to such decomposition, it is necessary to reconstruct the initial definition, to establish whether there may be inconsistencies or problems of some other kind in the rewritten definition. We thus cannot avoid John Locke's question – one mentioned at the outset of this chapter – of where at last we should stop when defining. It is worth repeating that we are obliged to stop somewhere. We neither stop because that is a fair solution (for it clearly is not), nor because resources available for this discussion are scarce. What forces the reliance on some set of primitive terms, or building blocks, are the exceptional difficulties that arise when relevant definitions of these primitives are sought. Illustrative in this respect is the term *being* and its related verb *to be*, for which Blaise Pascal, a seventeenth century French scientist and philosopher, summarizes the problem conveniently as follows:

On ne peut entreprendre de définir l'être sans tomber dans cette absurdité: car on ne peut définir un mot sans commencer par celui-ci, *c'est*, soit qu'on l'exprime ou qu'on le sous-entende. Donc pour définir l'être, il faudrait dire *c'est*, et ainsi employer le mot défini dans sa définition." (cited in Eco 1999)<sup>4</sup>

As Umberto Eco, a semiotician and writer, notes in his *Kant and the Platypus*, we can still say quite a bit about the verb *to be* while missing a definition. Along this same line of argument, it is thus relevant to elucidate the meaning intended for

---

<sup>4</sup> Author's translation: "We cannot attempt to define being while avoiding that paradox: we cannot define a word if we do not start its definition by *it is*, regardless of our saying or assuming this. It follows that, to define being, we should say *it is*, and thereby use the word we are defining within its own definition.

the primitives that are used. Any such primitive will be used as a building block of definitions, although itself will remain without a satisfactory definiens.

It would be naïve to think that primitive terms are used merely for reasons of efficiency. This would be the same as to consider that there are (or must be) definitions for all primitive terms, even if they are not defined within a given discussion. Even if we somehow manage to put aside Pascal's observation on the necessary circularity in some definitions, it is relevant to note the deficiencies that such an optimistic outlook subsumes. These continue to be reiterated at least since Plato's dialogues that seek a definition of, e.g., beauty in *Hippias Major*, justice in *The Republic*, or courage in *Laches*:

1. Suppose that the primitive term in some definiens is a property. In the absence of a definition of that property, we cannot tell with certainty whether something has that property. When in a conversation with Hippias, Socrates laments his failure to understand much in a precise manner and wonders "How will you know whose speech — or any other action — is finely presented or not, when you are ignorant of the fine?" (Plato 1997, *Hippias Major*, Sect. 304e). In other words, how would one know whether to call something fine if there is no definition of fine?
2. An undefined property cannot be related to other properties. In discussing the meaning of justice, Socrates says "For I know not what justice is, and therefore I am not likely to know whether it is or is not a virtue, nor can I say whether the just man is happy or unhappy" (Plato 1997, *The Republic*, Sect. 354b).
3. When there is no definition of a property, it is difficult to know what to advise to satisfy that property. In the dialogue *Laches*, two fathers seek advice on whether they should have their sons trained to fight in armor. Having received different opinions on the matter, they seek Socrates' advice. After establishing that the aim of the training is to instill virtue, and courage in particular to the children, Socrates goes on to question the various definitions of courage and virtue, and wonders "must we not first know the nature of virtue? For how can we advise any one about the best mode of attaining something of which we are wholly ignorant?" (Plato 1997, *Laches*, Sect. 190b–c)

While the three problems above persist undoubtedly for many terms, be they used as primitives in definitions or not, we evidently see that many are commonly used. While some degree of confusion in meaning is mostly admitted in communication, there are cases when precision is critical, yet absent. Such is the use of primitive terms in law for instance, where they are left undefined, often to the aim of remaining open to interpretation. Instead then of deciding on the precise meaning and consequently identifiable consequences of a chosen definition, legislation tends to leave the interpretation of some terms for their context of use. That is, a definition is not given, but must be made explicit within a particular setting, to which the legislation applies. This is for instance the case with the definition of torture in law. The early mentions of the term *torture* in law go back to the 1863 Lieber Code, which aimed to establish rules of war applicable to the US army. The Code uses torture as a primitive term and mentions it only once: "Military necessity does not admit of cruelty — that is, the infliction of suffering for the sake of suffering or for revenge,

nor of maiming or wounding except in fight, nor of torture to extort confessions” (Lieber 1863, Sect. 1, paragraph 16). Various definitions of torture have since been offered within international conventions and their local implementations, but few have generated as much controversy as the 2002 US Department of Justice memorandum, in which a later withdrawn interpretation was given of the term torture. The starting point for the memorandum was the United Nations’ 1987 *Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment*, the first article of which gives the following meaning to torture:

For the purposes of this Convention, the term ‘torture’ means any act by which severe pain or suffering, whether physical or mental, is intentionally inflicted on a person for such purposes as obtaining from him or a third person information or a confession, punishing him for an act he or a third person has committed or is suspected of having committed, or intimidating or coercing him or a third person, or for any reason based on discrimination of any kind, when such pain or suffering is inflicted by or at the instigation of or with the consent or acquiescence of a public official or other person acting in an official capacity. It does not include pain or suffering arising only from, inherent in or incidental to lawful sanctions.

It is not unusual to read vague terms in this and other definitions of torture, and in particular with regards to the degree or level of pain or suffering. That is, such primitive terms are found elsewhere as well; in an encyclopedia of philosophy, we find the following definition: “Torture is: (a) the intentional infliction of extreme physical suffering on some nonconsenting, defenseless person; (b) the intentional, substantial curtailment of the exercise of the person’s autonomy [achieved by means of (a)]; (c) in general, undertaken for the purpose of breaking the victim’s will” (Miller 2008). Both the term *severe* in the UN Convention against Torture and the term *extreme* in the last-mentioned definition are vague. It follows that all three issues that Plato raised will arise whenever we ask if some act was an act of torture. This is of course a recognized problem, the response to which is local interpretation, as in the 2002 US Department of Justice memorandum. The decision that would have stated more precisely what constitutes an act of torture is thereby delegated: the people making the definition leave some of it unclear, and thereby allow the user of the definition to choose how to complete the definition. As we shall discuss in more detail later on in this text, there is unlikely to be a stable and widely accepted definition of a vague term. Roughly speaking, while one individual may call severe some pain or suffering, another need not do so. Even if the two agree on what pain is severe, we can easily find a third who may disagree. The point is that there is no universal standard that says which level of pain or suffering is severe. To understand better the term *severe*, the authors of the Memorandum draw on dictionaries:

The dictionary defines ‘severe’ as ‘[u]nsparing in exaction, punishment, or censure’ or ‘[i]nfllicting discomfort or pain hard to endure; sharp; afflictive; distressing; violent; extreme; as *severe* pain, anguish, torture.’...Thus the adjective ‘severe’ conveys that the pain or suffering must be of such a high level of intensity that the pain is difficult for the subject to endure. (Bybee 2002, p. 5)

It is not difficult to see that the above quote does two services within the Memorandum. The first is a disservice, since it is clear to the careful reader that it only



manages to define the term *severe* with *difficult for the subject to endure*. This is a disservice, since it simply moves from one vague term (*severe*) to another (*difficult*), so that the gain in clarity and precision is small, or inexistent. The service it does is that such a paragraph gives to some the appearance that the authors did their best in seeking the ordinary meaning associated *severe*. Since doing so did not help much, their subsequent step was to look into the interpretation of the term *severe* in prior court judgments and other official use. From an available definition of *emergency medical condition* in a prior case, the authors of the Memorandum argue that *severe pain* “must rise to a similarly high level — the level that would ordinarily be associated with a sufficiently serious physical condition or injury such as death, organ failure, or serious impairment of body functions — in order to constitute torture” (Bybee 2002, p. 6). Despite its barbarism, this definition shows that the authors did place some restrictions on the meaning of *severe*. In particular, they increased the pain threshold, thereby reducing the variety of practices that are called torture. This still did not alleviate vagueness: there is the phrase *high level* in the last quote above. That reference to a *high level* has a very real counterpart in the interrogation practices. It reflects an unavoidable fact that different people react differently to same torture practices; to put it bluntly, some die faster. The vagueness that remains in the definition of *severe* in the Memorandum thus leaves some field of decision to physicians and psychologists who actively participate in violent interrogations. In doing so, the doctor plays the ethically unclear role of instruments that measure physical and mental pain, participating in acts that may amount to war crimes (Marks 2005). If the definition itself cannot say what that level is high, someone must decide what this means during interrogation.

While it may not be possible to avoid having primitive terms in a definition, which elude precise and agreed upon meaning, it is important to remember that a definition always serves specific purposes. The aims, especially in cases as that of the Memorandum, go well beyond the ideals of mutual understanding and clarification. Once we admit it impossible to draw a sharp line between severe and nonsevere pain or suffering, the problem is no longer to seek a final definiens for a vague term, but to study the purposes, to which a particular interpretation thereof is made. As a lawyer illustriously observed (Omer Ze’ev Bekerman 2005), any act that falls short of blunt and effective sadism is not banned by the Memorandum. In a context of the so-called Global War On Terror after the events on 11 September 2001 across the US, the definition laid out in the Memorandum simply reflected the official positions. These are best summarized by two quotes, one from respectively the then US president and vice-president. The former famously proclaimed “[e]ither you are with us, or you are with the terrorists” (Bush Sept. 20, 2001), while the latter was quoted suggesting that the terrorists “don’t deserve the same guarantees and safeguards that would be used for an American citizen going through the normal judicial process” (Slevin and George Larnier Nov. 15, 2001). We noted in Chap. 2 that myths and taboos in tribal societies place restrictions on an individual’s field of personal decision. Statements given by figures of authority have a similar effect in any form of government. Any seemingly rigorous discussion of the primitive terms in a definition of torture happens within such bounds. We can, given the official



statements and the necessity to protect abusive interrogation practices, foresee then the direction in which the primitive terms will be interpreted.

When confronted to primitive terms, it is crucial to read a given definiens in relation to the purposes that it is called to serve. More generally, any reading will be more informative if it considers not only the definition of interest, and the primitive terms therein, but also the cues (communicated verbally or otherwise) to the beliefs, desires, intentions, and attitudes of those using the definition. In order to account for these considerations, we will consider later in this section the definitions thereof (i.e., of desires, beliefs, intentions, and attitudes), and put these definitions to use throughout the rest of the text.

### 3.1.3 *Criteria from Standard Theory*

The previous discussion of definitions led us to make a distinction between those of ostensive and intensional kinds, then discuss what role and consequence is brought using primitive terms. Although this showed that definitions may be deficient, and that deficiency is very hard to avoid in some cases (as when some primitive terms are vague), it is still relevant to ask if and how we can try to design definitions that we could call rigorous, despite the said shortcomings. The term *rigorous* is itself vague. Instead of asking how rigorous a definition should be to be qualified as rigorous, it is better to ask this differently: What criteria should a definition satisfy in order to be called rigorous?

It is unlikely that the questioning of specific definitions (say, of courage or justice) has gone without the study of criteria for the evaluation of qualities we may hope these definitions to have. The so-called standard theory suggests two criteria, called *eliminability* and *conservativeness*. This standard account has been articulated throughout the twentieth century, starting within the Lvov-Warsaw School of Polish philosophy. We will go over the two parts of the standard theory below, and in doing so follow Nuel Belnap, a philosopher (Belnap 1993).

A definition satisfies the *eliminability* criterion if and only if it satisfies the following: in any context, in which we encounter the definiendum, and replace it with the definiens, the meaning of the context does not change. If we were to replace the definiendum with the definiens, this would not change what we understand from the context, in which this replacement was effected. There are several obvious difficulties with taking this criterion in the form that was just offered. Namely, there are the terms *context* and *meaning*, which require clarification. We may encounter a term within some written text, so that the text itself may be considered as the context; if we instead hear it in a conversation, that particular conversation could be its context; and so on. A definition that would satisfy the stated eliminability criterion would be a universal definition, that is, one that is independent of the context of use. Clearly, such definitions are at best rare and much more likely elusive. If you take a term as straightforward as *apple*, you can ask whether it carries the exact same meaning when used with or without quotes, in Christian mythology or an economic

description of a fruit market, and so on. The obvious answer is that it does not. We are consequently likely to be asking a definition to satisfy the eliminability criterion within some context, for which it is tailored. Thus, we will have different definitions for the term *apple* in Christian mythology and in botany, while both of these definitions may be designed so as to satisfy the eliminability criterion *within* their context of application.

What about the term *meaning* in the eliminability criterion? Suppose that we are given some context, say a paragraph from a text, and that we replace each occurrence of the definiendum with the definiens in that paragraph. If we put side by side the paragraph with definiendum only and the paragraph with definiens only, we can compare them. The question then is what will we compare to establish whether the definition that related the definiendum to the definiens satisfies the eliminability criterion? Belnap's answer is that we can compare the conclusions we can draw from the two paragraphs; if they both lead us to the same conclusions, then the definition satisfies the eliminability criterion. It is of course clear that the replacement should be such that the new text has no occurrence of the replaced term. In other words, the definition must not be circular: the definiendum must not appear in the definiens. More generally, we can say that the criterion of eliminability requires the following:

1. *Noncircularity*: the definiendum is not contained in the definiens.
2. *No inferential enrichment*: within the context  $C$ , in which the definition is intended to be used, if every occurrence of the definiendum is replaced by the definiens, thereby giving us a context  $C'$ , same conclusions can be drawn from  $C$  and  $C'$ .

The criterion of *conservativeness* requires that a definiens explains the meaning intended for the definiendum within a given context, and does *only* that. A definiens that adds information beyond what is necessary to explain meaning is not conservative, if that additional information allows us to draw conclusions, which we could not draw before the definition was available. It is unimportant if these additional information or conclusions are of interest. They may be, but their place is not in the definiens. For a caricatural example, we can take any definition of torture and add to it recommendations on what practices may be more or less effective; such additional information may be of use to an inexperienced sadist, but is an undesirable surplus in the definition. Why is this so? The use of the nonconservative definition in reasoning may lead to contradictory conclusions. In such a case, the definition evidently no longer serves its purposes. If the definition without the added information, that is, the conservative definition does not produce contradictions, then the additional information is certainly an undesirable surplus.

The eliminability and conservativeness criteria suggest guidelines to follow when designing definitions. How to aim for noncircularity is clear: avoid repeating the definiendum in the definiens, or more generally, avoid having synonyms of the definiendum in the definiens. This will clearly be violated in some cases, as in attempts to define some primitive terms; it is useful to keep in mind Pascal's observation that some circularity seems unavoidable, such as when the verb *to be* is defined.

There will be inferential enrichment if the relationship that the definiens gives involves terms (both primitive or otherwise) that are absent from the context, in which the definition is given. To avoid inferential enrichment, the definiens should establish a relationship between the primitive terms that have *already* been introduced in the context. The context itself can be almost anything spoken, written, or otherwise communicated and/or made available. Regardless of how it is made available, the context will be the backdrop against which a definition is given. The rule to follow against inferential enrichment is to introduce any term used in the definiens before providing the definition. The context should be outlined first, along with definitions of any terms that should prove useful in the definition.

A final, third rule is required to make the definition conservative. Consider a caricatural example of a cynic who sets out to define the term *debt*. A convenient candidate for such a definition is in *The Cynic's Word Book*, from Ambrose Bierce, a late nineteenth and early twentieth century journalist and satirist: debt, it says, is "an ingenious substitute for the chain and whip of the slave-driver" (Bierce 1906). While this may be an unusual way of conceiving debt, the definition satisfies the noncircularity criterion. If placed in a different, noncynical context, such as a lending contract, this definition would fail the inferential enrichment requirement: the reader of the contract could unexpectedly conclude that the lender is in fact a slave-driver. However, if within a proper context, such as a cynical text about lending practices, then the definition need not lead to inferential enrichment. Regardless of the context, however, the cynic's definition is not conservative, because it uses an embellishment that is not absolutely necessary to grasp the meaning that the cynic seems to intend for the term debt. If the word *ingenious* is removed, debt is still a substitute for the chain and whip of the slave-driver, that is, not much is lost. Whether more can be removed or rewritten depends on the intentions of the individual designing the definition. The rule to meet the requirement of conservativeness is then the following: remove all words and rewrite all terms in a more concise manner; do so as long as the context and the definiens lead to the same conclusions as those that were allowed by the initial, nonconservative definition. In other words, conservativeness requires economy of expression.

Following one of the guidelines without taking care to follow the others can produce deficient definitions. In the course of redesigning the cynic's definition of debt to make it conservative, it is possible to fall into the trap of inferential enrichment. For example, if the phrase *a substitute for the chain and whip of the slave-driver* is replaced by the more concise definiens, the phrase *a substitute for any tool of oppression of the slave-driver*, then *debt is a substitute for any tool of oppression of the slave-driver* could be offered as a cynical definition of debt. The latter definiens (i.e., the phrase *a substitute for any tool of oppression of the slave-driver*) is more general than the former definiens (i.e., *a substitute for the chain and whip of the slave-driver*), since it says that debt is a substitute for *any* tool of the slave-driver. If the intention of the cynic was to say that debt is a substitute of *only* the chain and whip, then the new definiens does more than the initial one, so that the definition that would use the new definiens produces inferential enrichment.

It seems that the criteria and the ensuing rules allow the individual to cut through the bad definitions to get to the good ones. The criteria have some merit, certainly, but mostly as something to be aware of, and not, as hoped, of something to diligently apply and expect good definitions. The problem is that noncircularity and inferential enrichment can never be satisfied for definitions given in a natural language, such as English. Roughly, they work as long as not many questions are asked. What would such questions be? Start with noncircularity. To have a circular definition might appear to some as an attempt to fool or make fun, for how can something be defined via itself? Still, we see by the very construction of dictionaries that circularity can hardly be avoided altogether. To put the blame on the makers of dictionaries would be to assume that at least some terms require no definition at all, or equivalently that they are so clear to everyone that no explicit meaning should be given to them. This is unconvincing, given how much debate has surrounded notions of, say, justice and equality throughout history. Move then to inferential enrichment. Inferential enrichment happens within a given context, whereby the context is the text or conversation containing the definition. We can only ensure that there is no inferential enrichment if the context is fully specified and understood in exactly the same way by anyone involved in the communication that occurs through that context, and only if the individual performs inference through deduction via an explicit set of reasoning rules. Basically, inferential enrichment can only be ensured in a mathematical logic, in which both the context and the reasoning rules must be made explicit to perform any reasoning in the first place. The definition of torture will be interpreted differently in different contexts and by different people. Hence, some may conclude that an act of violence is torture, while others will not. Apart from laying the blame on one or other definition of torture, which fails the inferential enrichment criterion, we should rather wonder whether it is possible at all to design a definition of torture in plain English that will not fail this same criterion. Some of the very best attempts to a definition of torture (Miller 2008) still include vague terms, and thereby let inferential enrichment happen. Far from being somehow inappropriate, doing so seems elusive. These limits on criteria for rigorous definitions need not induce pessimism, but they should neither inspire awe.

## 3.2 Signs to Objects and Back, via Concepts

A definition amounts to a series of signs organized according to the conventions of the language in which the definition is written. Familiarity with the language allows the recipient of the definition to distinguish letters, numbers, punctuation marks, white spaces, and other signs that may appear therein (e.g., signs for integrals, sums, etc.), so as to identify words, phrases, and other combinations of signs that form larger fragments of communicated information. Based on the conventions learned through experience, the recipient can claim to understand the definition, that is, to know what the combined signs refer to, what relationships they convey between that which they refer to, and how these references and relationships can be used in

subsequent reasoning. The purpose of this section is to clarify the ideas laid out in the preceding sentence.

The observations made in the paragraph above are equally valid if we replace the term *definition* with *advice*. As advice is a tool of coordination, and the latter cannot happen without communication, any advice manifests itself as a collection of signs organized according to some rules.

### 3.2.1 Signs

The first task is to clarify the term *sign*. This is a primitive term in semiotics, which “is concerned with everything that can be taken as a sign” (Eco 1976, p. 7). It is convenient to take a sign to be anything that can be communicated, regardless of the medium. In defining sign in this way, this term no longer carries the burden of being a primitive term. Instead, *communication* becomes the primitive term, and we do not go into its definition. If we are looking for an ostensive definition of *sign*, then any letter, mathematical sign, line, point, sound, and so on can be pointed to and called a sign. Why mention the medium of communication? While some signs will be transferred from the source to a destination via speech or text, others may require the predominant intervention of senses other than hearing and vision, such as touch. What senses may be involved, or what medium of communication privileged will determine the kinds of signs that can be communicated. The medium becomes important once we recognize that not every sign can be translated into – is not equivalent to – another sign. This merely says that it is dubious to claim that any message can be conveyed to a same effect through any medium.

Our concern is with defining advice, in addition to studying definitions themselves, both as the potential content of advice and as something that is assumed when advice is dispensed. Although advice and definitions may be dispensed via various media, it would overextend the present discussion if we attempt to encompass any communication medium, and thereby any potential sign. Since what you are reading here is written text, we are limited only to some signs, namely those available when writing. The very practical consequences of this remark are the following. When a definition or advice is reproduced as-is from another source, where it is also written, signs are reproduced truthfully (i.e., they are copied). Convention indeed commands that citations be placed in quotes and any changes thereto explained. However, when we give examples of spoken advice or definitions, we are translating the signs used in that communication, so that any analysis thereof is bound to do some injustice to the individual or machine having originated that communication. When Pericles’ funeral address was quoted in Chap. 2, it was reproduced as-is from the sources we cited, so that it is truthful to those sources. However, we cannot claim the same for Thucydides’ transcription of Pericle’s address. Maybe Pericles dispensed some of his advice with irony or sarcasm, and Thucydides simply did not notice when writing from memory Pericles’ address. We cannot know. Of course, that we cannot know does not necessarily mean paralysis; relevant discussion can

be given from the signs we have, even though it will necessarily be limited. Here, what we have are written signs only, even when they are claimed to represent spoken or otherwise communicated advice or definitions. To understand this is to grasp that we have necessarily lost some signs used in communication, and that we are working only on what could, within the limits of attention and signs available in the language, be translated into signs that are used within this text. Some of what will be lost in translation from speech to text are forms of nonverbal communication, which Mark Knapp and Judith Hall identify in their *Nonverbal Communication in Human Interaction* (Knapp and Hall 1978), such as the impact of environmental factors on communication (e.g., furniture, architectural style, lighting, smells, colors, temperature, environmental noises), of proxemics (e.g., spatial arrangements of seating or standing, conversational distance, orientation), of kinesics (body motions, such as gestures, facial expressions, eye movement), of physical contact (touching, hitting, greetings, farewells), of physical characteristics (physique, height, weight, hair, skin color), of paralanguage (voice pitch, volume, tempo, silences, speech errors), and of artifacts (objects manipulated by the individuals as communication occurs). The best we can do within the bounds of written text is describe some of these in the text itself, but much will clearly be lost since it is evident that experiencing communication and reading about it after it had happened are very different.

Having delimited the range of signs of interest, recall that any definition and any advice is merely a collection of signs. Any evaluation of an existing definition, just as the design of any new definition, cannot be performed by the simple manipulation of signs. The signs that end up being used to write down or say a definition are merely the result of the act of communicating the definition. They are the product of that communication, which makes the definition available to recipients precisely by way of signs. This simple observation has obvious, but important implications for the design and evaluation of definitions. To verify whether a definition satisfies the eliminability and conservativeness criteria requires the interpretation of the signs, or in other words that meaning be read from the signs. An exception to this should be mentioned, although it will not be revisited further below, namely, in those natural languages, in which there are no declinations, it is still possible to detect circularity even if the language is unfamiliar. Visual cues are in that case enough to spot patterns, that is, to see that the signs in the definiendum appear also in the definiens. However, the evaluation in all other cases – more numerous without doubt – requires that the term *meaning* be clarified. The exception aside, the evaluation of inferential enrichment and conservativeness in a definition cannot be done by simply looking at patterns of signs therein. Same applies for the analysis of advice. The individual who is not a speaker of Hindi will not be able to analyze either definitions or advice communicated in signs that are organized according to the rules of that language.

### 3.2.2 *Objects and Concepts*

What seems fairly uncontroversial is that every sign refers to something. It is in this relation, that is, of a sign standing for, or equivalently referring to something

else that meaning is created. That is, there is no meaning without the reference relation. When an ostensive definition is given, the reference relation is rather explicitly stated. For example, the thing that the sign stands for can be pointed with a finger, and the act itself establishes the reference relation as long as there is a convention between all those present that such an act is intended to establish such a relation. The act of defining establishes precisely that reference relation, although it can only do so via the manipulation of signs alone, and not of things that these various signs refer to. In simpler terms, a definition will relate some signs to others, while assuming that the reader of the definition already knows what the latter ones refer to.

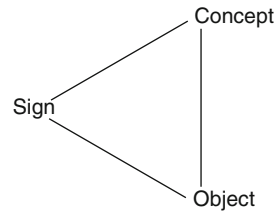
In addition to the sign and the thing it represents, it has been recognized at least implicitly and since Aristotle's *De interpretatione* that something more is necessary. As Eco argues in an essay (Eco 1987), Aristotle designed what is today usually called the *semiotic triangle* in the Anglosaxon tradition of linguistics and semiotics, that is, a ternary relation "in which words are related on one side to concepts. . . and on the other to things" (Eco 1987, p. 552). We can equate, in that quote, the terms *words* to our *signs*, and *things* to whatever it is that the signs refer to; but what is a *concept* then? An answer lies in the distinction that Gottlob Frege, an eighteenth and early nineteenth century mathematician and philosopher, suggested between the terms *object* and *concept*:

The concept (as I understand the word) is predicative. [His footnote:] It is, in fact, the reference of a grammatical predicate. [End of his footnote.] On the other hand, a name of an object, a proper name, is quite incapable of being used as a grammatical predicate. This admittedly needs elucidation, otherwise it might appear false. Surely one can just as well assert of a thing that it is Alexander the Great, or is the number four, or is the planet Venus, as that it is green or is a mammal? If anybody thinks this, he is not distinguishing the usages of the word 'is'. In the last two examples, it serves as a copula, as a mere verbal sign of predication. . . In such a case we say that something falls under a concept, and the grammatical predicate holds for this concept. In the first three examples, on the other hand, 'is' is used like the 'equals' sign in arithmetic, to express an equation. [His footnote:] I use the word 'equal' and the symbol '=' in the same sense as 'the same as', 'no other than', 'identical with'. . . [End of his footnote.] In the sentence 'The morning star is Venus', we have two proper names, 'morning star' and 'Venus', for the same object. In the sentence 'the morning star is a planet' we have a proper name, 'the morning star', and a concept-word, 'planet'. (Frege et al. 1951, pp. 169–170)

In the quote above, "a name of an object" equates in our terminology here with the phrase "sign of an object" (i.e., the term *name* equals the term *sign*). Following the same line of reasoning as in the quote, the sentence "John is a mathematician" can be rephrased as saying that a person referred to as John has the property of being a mathematician. The *object* in the sentence is that which is referred to by the sign "John," while the sign "is a philosopher" refers to the *concept* of being a philosopher. There are some parallels to be made here with the distinction between ostensive and intensional definitions; the former define by identifying objects, while the latter define by giving the properties that should be satisfied by objects. An intensional definition thus identifies objects indirectly, by pointing to a concept; objects that are concerned then by that definition are all those that an individual using the definition perceives as having the properties that the concept predicates. Note in passing that,



**Fig. 3.1** Sign/concept/object triangle



while we can have one concept per property, we can also have a concept from many properties; there is no requirement that a concept arises out of a single property.

### 3.2.3 *Relating Signs, Objects, and Concepts*

What are called *concepts* here mediate between the signs used in language, and the objects these signs refer to. Figure 3.1 puts lines between the terms *sign*, *object*, and *concept* to refer to the relations they stand in. Drawing the lines as shown in the figure is rather conventional within semiotics. As Charles Peirce, a nineteenth and early twentieth century philosopher and semiotician suggested, “[a] Sign, or Representamen, is a First which stands in such a genuine triadic relation to a Second, called its Object, as to be capable of determining a Third, called its Interpretant, to assume the same triadic relation to its Object in which it stands itself to the same Object” (Peirce 1998, p. 272). As a side note on terminology, our *sign* is his Sign (equivalently, Representamen), our *object* is his Object, and our *concept* is his Interpretant, what he elsewhere called the mental effect on the individual.<sup>5</sup> While we have clarified to some extent the three terms that Peirce and others, both before and after him, have identified and related, we have not said much of what these drawn lines, themselves signs, may stand for.

In relating the three, Peirce argued for a ternary relationship, saying it “is genuine, that is its three members are bound together by it in a way that does not consist in any complexus of dyadic relations.” (Peirce 1998, pp. 272–273). A way to attack this claim is to consider whether we can offer any case, in which there is a sign and a concept, and no object. For instance, the written word *Hercules* may be related to a concept of having the character traits that some literature may call Herculean. This concept would be an aggregate of properties, some of which would to act bravely in danger. What would be the object here? Or, in other words, is there an individual who was ostensibly defined as Hercules? Equivalently, is there a person whose mother pointed his finger at him at birth, and in presence of others spoke the word Hercules, or in more recent times, wrote that term down in some official book that

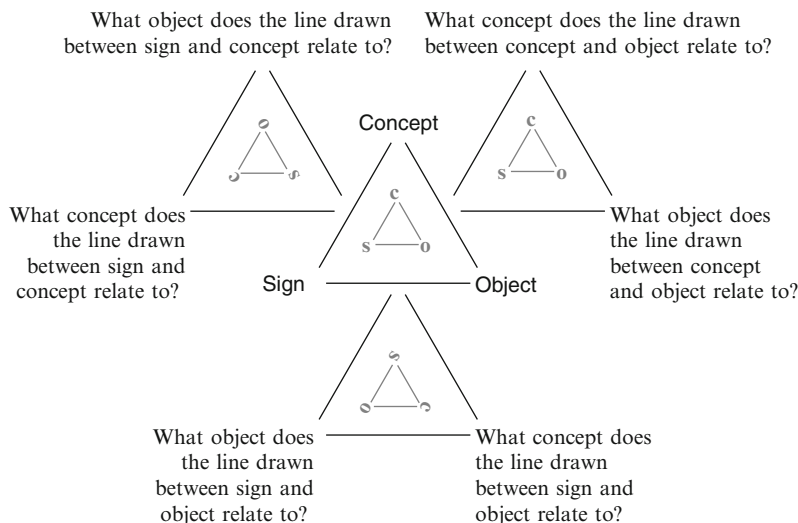
<sup>5</sup> “A representation is that character of a thing by virtue of which, for the production of a certain mental effect, it may stand in place of another thing. The thing having this character I term a representamen, the mental effect, or thought, its interpretant, the thing for which it stands, its object” (Peirce 1932, Vol. 1, p. 564).



holds the names of citizens? Since parents can be inventive, there are probably people who are living or have lived, and who were named Hercules. In this sense, there exists or have existed objects referred to by the sign Hercules. However, is there a ternary relationship between someone named Hercules at birth, and the object(s) intensionally defined by the concept (i.e., the properties aggregated by that concept) that the sign may evoke to an individual versed in Greek mythology? To answer the affirmative would be to mistake some particular (maybe a coward) Hercules who lived, and the mythical individual or creature called Hercules. To the extent that there has been no object having the properties that myths attribute to Hercules, there is no object in the ternary relationship. Does this mean the failure of there being a ternary relationship for *any* sign, concept, and object? We can straightforwardly see from events and situations that can be empirically observed that there is often this ternary relation, but that the object can also be missing from the discussion, in which the sign and concept are evoked. We can thus look at the question in another way and provide the following answer: if all three parts of the relation are mentioned, then we can speak of a ternary relationship; instead, if only two are mentioned, we can still say something about the binary relationship that stands between each pair. The question is consequently no longer whether the binary relationships are alternative to the ternary one, but what to call them when some element of the triad is simply not considered at all. The point here is that we can, without committing serious fallacies, still talk about the binary relations in the semiotic triangle, but also acknowledge that doing may leave open any question raised on the relationship of those elements of interest to others that may not be considered.

With the above disclaimer in mind, the task is to describe the binary relations between sign, concept, and object. A way to approach this task is to recognize that each of the three lines drawn in Fig. 3.1 is a sign. Each of the lines can be put in place of a sign in the sign/concept/object triangle, which leads to two questions for each of the lines. What object does a line relate to in its sign/concept/object triangle? What concept relates to the line? Where these questions come from can be easily pictured from the contents of Fig. 3.1, as shown in Fig. 3.2. The latter includes, at places where we would expect objects and concepts, the variants of the two questions above, and which need to be answered for each line drawn in Fig. 3.1.

Before we actually consider the questions in Fig. 3.2, we should ask first whether it makes any sense to manipulate the sign/concept/object triangle in the way shown in that figure – to have three of them around the basic one from Fig. 3.1, each having in place of the sign, the line drawn between any two of *sign*, *concept*, and *object* terms from Fig. 3.1. If not, then the questions in that figure make no sense. If we want an affirmative answer instead, and since each thin black line in Fig. 3.1 is used as a sign for a relation, we should say whether a relation is an object. It is easy to argue that a relation is an object, not because of the relation itself, but because of the very definition of an object. Although some may read an object as being something that must have some specific properties, such as being tangible, this is not intended with the term *object*. Peirce argued that the semiotic triangle has general applicability: we can have anything as a sign (e.g., smoke can be a sign of fire, where smoke is a sign and fire is the object), and we can have anything as the object



**Fig. 3.2** Seeing the relations in the sign/concept/object triangle as signs, and the questions that doing so gives rise to. The figure was drawn by taking the basic sign/concept/object triangle from Fig. 3.1, then adding three triangles and rotating them so that the sign corner is turned to a line drawn in the basic sign/concept/object triangle. The concept and object corners of each of the three triangles give a question that requires an answer to characterize the relations between sign, concept, and object terms. The corners of the small s/c/o triangle highlights the orientation of the sign, concept, and object corners of the triangle, inside which the small s/c/o triangle is drawn

(e.g., drawing of smoke is a sign, and smoke is in that case the object). That pretty much anything can be a sign and anything can be an object should not confuse: namely, that a written word *horse* is a sign does not make it run like a horse, that is, along the common wisdom, the word is not the thing. A particular sign is usually not at that same time the object it stands for. If we continue with the above examples: smoke is not fire, and the word *horse* is not a living horse. Hence, Fig. 3.2 is not nonsensical, and the questions asked therein are relevant.

To affirm that a relation is an object is not terribly interesting: since an object can be anything, we still know little of what a relation may be. Objects called properties tend to raise that the same philosophical issues as relations (Swoyer 2008). Relations are thus typically taken as a kind of properties, and the arguments that can be made for properties in general stand for the particular case of relations. What are these other properties, which are not relations? When a tree is described by someone uninterested in plants, properties will be given such as size, age, and color. When age is given, there seems to be no other object with regards to which age is estimated. That is, while we may say that the tree is left of the house, thereby relating the tree and the house, we seem not to be relating when we speak of color, age, weight, and so on. There is consequently a convention, to distinguish relations, or relational properties (as properties involving more than one object) from monadic properties; monadic, because they are properties of (or involving) a single object. While we

may settle on this convention here, to do so is more a matter of convenience than a reflection of a consensus. Consider for instance weight. Roughly speaking, it relates the object being weighted to planet Earth (or Moon if you are weighting something on that celestial body instead). Specific weight, speed, mass, and force may be considered for some practical purposes as monadic properties, but they are always experienced within a frame of reference (or context), so that they necessarily relate the object to some other object, and effectively seem to be relational. Color, which at first sight seems rather nonrelational, is anything but: color is, so to speak, in the eye of the beholder, that is, of the individual perceiving the reflection of light from a particular surface. We see here again that the use of words should be taken more seriously only after accounting for the wider setting, or context of their use.

There seems to be no solid argument that settles the question of whether the distinction between monadic and relational properties is a valid one, that is, if these two are truly distinct objects, and not just names for the same objects. What can be safely affirmed, however, is that this distinction is useful mainly as a means to hide some detail, which an individual takes as irrelevant to his intentions. Even if someone was to somehow conclusively demonstrate that all properties are relational, there would still be a use for monadic properties, since these would simply be those relational properties, for which only one of the relata (i.e., object participating in the relation) would be of interest to the intentions of the individual using that property.

Having distinguished monadic and relational properties, we can provide the first part of the answer to three of the six questions in Fig. 3.2. Namely, a relational property is the object for which (a) the line drawn between *sign* and *concept* relates to; (b) the line drawn between *sign* and *object* relates to; and (c) the line drawn between *concept* and *object* relates to. This is only part of the answer, however, since these lines do not relate to *any* relational property (e.g., they do not relate to *being left of*, or *being smaller than*).

Meaning seems to be created from the reference relationship that stands between the sign and its object. We said that a sign refers to, or stands for an object. We can call the relational property between *sign* and *object* the *reference relation*, and use this relation in text by saying, as needed, that a sign references an object, or that an object is referenced by the sign. These are mere considerations about the sign that can be used to refer to the reference relation itself, while describing the reference relation requires an account of the mechanism of reference, the mechanism by which people relate signs to objects. These are roughly the main competing descriptions of such a mechanism:

- The descriptive mechanism of reference, due originally to Frege and Bertrand Russell, a late nineteenth and twentieth century philosopher, says that reference operates by matching somehow the object to the properties, i.e., a concept that an individual has in mind. A sign then fits the object only if the individual relates to the concept in question. If an individual relates some set of properties, such as some range of sizes, weights, colors, forms, and behaviors to the sign *horse*, any object that falls within these ranges and across exactly these properties will be referred to by the sign *horse* by that person. If we were to speak with an individual who behaves along this mechanism of reference, we would need to

hypothesize or somehow discover his concepts before we can understand what object his signs refer to. We could find this out by asking him to intensionally define the terms he is using.

- The descriptive mechanism seems to be rather demanding on the individual: to have a reference between a sign and an object, an individual is expected to have a concept that uniquely identifies that object. In order to call an object by the sign *horse*, the concept should involve such set of properties and such constraints on the values of those properties that we cannot be lead to refer by *horse* to an object typically called a zebra. That is, the description that the concept gives is, so to speak perfectly precise. Clearly though, people tend to refer to objects they know little about and still make meaningful conversation. This is exactly what is going on as you are reading this text: it is rather idealistic to claim that both I and you have a perfectly clear idea of what each of the words you read on these pages refers to. To avoid this, and some other issues, Saul Kripke, a contemporary philosopher, tells us that the reference of a sign to an object is established by some initial act of naming, a sort of baptism of the object. His argument is for a causal mechanism of reference. Any use of the sign after that initial naming can be related to the initial naming via a causal chain of events. The reason an individual calls some object a *horse* is thus because there has been an initial act of naming some object with the term *horse*, and there is a transmission (via, e.g., word of mouth, books, etc.) of that reference relation over time, people, places, and others; the individual in question happened to participate in that transmission at some point (e.g., by learning the names of animals from a picture book) and thereby learned to refer to an object by the term *horse*.
- An analogous problem to that raised against the descriptive mechanism can be offered to attack the causal mechanism: the causal mechanism is again very demanding, both on the individual and on the transfer relation. The reference between an object and its sign should be passed on from one individual to the next, and no errors should arise in that transfer. Another problem is that the causal theory has trouble with reference change, for it does not say why it happens that what we today call America was Christopher Columbus, the Genoese navigator and explorer called India, or indigenous peoples there called by other names. Another problem, which causes trouble both to the descriptive and causal mechanisms is the use of indexical terms, such as *I*, *you*, *now*, *here*, *far away*, etc. While these terms do refer to objects, individuals use them to refer to different objects at different times and places: *here* may refer to place  $x$  at time  $t_1$ , but then be used again to refer to another place  $y$  at another time  $t_2$ . Since the places are different, that is, have different properties and, or values for the properties, the descriptive mechanism fails: the properties and their values need not play no role in referring an indexical to an object. If an individual hears another baptizing place  $x$  at time  $t_1$  via the term *here*, the causal theory would have the hearer reuse the term *here* to still refer to the pair  $(x, t_1)$  even though he is at some other time and place. Since *here* is barely ever used to refer always to the same place, the causal mechanism fails as well with indexicals. Response to these troubles comes in giving a more important role to the characteristics of the context, in

which signs, including indexical are used. Space and time, elements of nonverbal communication, considerations of culture and social conventions act as cues to those involved in the context, in which the sign is used, and intervene more or less importantly in telling what object the term refers to. There are various such contextual mechanisms of reference (McGinn 1981; Wettstein 1984), which emphasise some elements of context over others. Together, they are maybe not as clean and simple as the descriptive and causal mechanism of reference, but are significant in underlining that the various conditions within which communication happens do affect the formation of reference relations between signs and objects.

If the reference relation forms between signs and objects under the influence of the context, the term *context (of reference)* should be clarified. We thus arrive to the term *context* again here, as we did when discussing the eliminability and conservativeness criteria for rigorous definitions. We were content there to take the context of a definiendum as the conversation or written text, in which we encounter that definiendum. That we encounter the term *context* again is not unexpected, since the discussions of definitions and those of reference cannot sharply be separated: the definiens for any definiendum arises from the reference relation. The definiens for the term *horse* will refer to some objects in case the definition is ostensive; if the definition is intensional, the definiendum will refer to the properties and their values, which form the concept related by the sign/concept/object triangle to the term *horse*. No definition can be designed without the reference relation; no advice can be analyzed without it.

To acknowledge that reference is established *in a context* is to see in a very different way the sign/concept/object triangle in Fig. 3.1 and those in Fig. 3.2. What we perceive there as the white space, on which these triangles are drawn, is not empty. Equating its whiteness with emptiness is misleading. The context in which the triangles are drawn contains what was judged necessary to obtain an interpretation of the signs in that triangle (the black lines, the words written next to the lines). The context is this very text, along with what stands beyond the text itself, that is, the background that the individual reader brings when reading, and the conditions in which such reading takes place.

The context includes anything that hearer or reader is normally aware of over the course of the communication, in which the sign of interest is mentioned. We can reconstitute at least some of what he is aware of by combining whatever is needed for any of the mechanisms of reference to function. In doing so, we venture that none of the mechanisms is good enough to explain the formation of the reference relation in *all* cases, but that each of them highlights a way that may provide only part of the explanation for the referencing of a particular sign to a specific object:

1. For the descriptive mechanism to operate, the individual should be able to perceive some properties. The context should be such as to allow this perception to happen, so that it must include visual, auditory, olfactory, or other cues.
2. The causal mechanism tells us that some terms can only be referred to objects if the causal chain of events, from the initial baptism, is known. The individual can

thus be aware of the prior uses of the term; that is, his knowledge of prior uses will be part of the context.

3. Philosophers have argued that individuals take various criteria into account to choose what object a sign refers to. The purpose of such criteria is to cut through the noise of the context to identify the object that best satisfies the criteria. Christopher Gauker, a philosopher, suggests a list of such criteria, which he argues guide the hearer or reader in relating an object to a pure demonstrative. While he does restrict himself to pure demonstratives (e.g., *this*, *that*), which are one kind of indexical [another being the so-called pure indexicals (van der Auwera 1980), such as the first person singular *I*] the criteria he offers still apply to most, if not all other writable signs that we have developed an interest in here. This is because pure determinants are likely to be the most difficult case, which requires considerable information from the context for the reference relation to be established. Gauker's accessibility criteria are the following (Gauker 2008):

- (a) **Salience:** The object to which the sign refers should be something that the hearer or reader can easily see, hear, or otherwise sense. This does not mean that the object must be the one in the context to draw most of the attention. Also, salience is taken in a broad sense and may involve visual, or auditory, or olfactory, or any combination of these ways, in which the object can be identified.
- (b) **Prior reference:** The object should have been mentioned implicitly or explicitly before in the current context.
- (c) **Relevance:** The object should be something that those involved in communication perceive as being of interest to the goals of that communication.
- (d) **Charity:** The sentences containing a sign should be perceived as "reasonable" by the hearer or reader when that sign is referred to the object. Suppose the first sentence of a definition, given in a sports manual, is this: *soccer is a team sport played by two teams of 11 players each, using a spherical ball*. What the charity criterion says is that if the hearer or reader refers the term *spherical ball* to a spherical ball made of concrete, then the sentence italicized above says something unreasonable, provided that the reader or hearer has seen a game of soccer, or played in one.
- (e) **Pointing:** If the speaker points to (by nodding, pointing the finger, or glancing in the direction of, or does whatever else is usually understood as pointing to) an object, then the object being referred to intersects with the imaginary line that the pointing act produces. In some cases, it is not the object pointed to that will be the referred object, but possibly the object it is part of, or is usually associated with.
- (f) **Location in a series:** When objects are ordered in a series, *this* or *that* can refer to the next item that comes to the attention of the individual observing that series:

Suppose we work on a candy bar assembly line as quality inspectors. We never think there is anything wrong with any of the candy bars, but we think that, in order to keep our jobs, we had better declare a candy bar defective now and then and pull it from the assembly line. Without even looking at the candy bars passing before me,

I declare, at appropriate intervals, 'That's good', 'That's good', and once in a while, 'That's bad'. Whenever I say 'That's bad,' you have to pull the passing candy bar from the assembly line. So in each case, my demonstrative refers to the candy bar passing before you. (Gauker 2008, p. 365)

The criteria above place constraints on their context of application, or equivalently, the context of reference. Saliency is not unlike what is needed for the descriptive mechanism of reference to work. There is further some similarity between prior reference and the causal mechanism, although such comparisons should not be crude: the causal mechanism requires that we can trace back the use to an initial baptism, while the prior use is merely a question of historical use and requires no causal chain of events to exist. The charity criterion is a different matter, involving a vague notion of reasonableness of a sentence, in which the referred term is used.

4. Context cannot exclude the verbal and nonverbal communication, which mentions the sign of interest in the first place. However, is there more in the context that would come from that other person? Suppose that the hearer can read minds along the lines of, e.g., the alien intelligence able to materialize the thoughts of astronauts, in Andrei Tarkovsky's 1972 science-fiction movie *Solyaris*. Now, if we leave out the materializing part, what remains is an ability to somehow see into the thoughts of the speaker or writer. If that was possible, then the reference relation might also form under the influence of the result of mind-reading. The hearer would know if what was communicated corresponds to what was intended, desired, believed, or otherwise by the speaker or writer. The context could thus include such considerations as well. Even if we reject mind-reading, it is still possible to argue that the intentions of the speaker or writer can be concluded from his verbal and nonverbal communication. The act of communication itself requires intentions: an individual will intend to use some sign in communication to refer to some object only if he believes that he can succeed, using that sign, to refer to that object in a given context of reference. It could thus be argued that the reference relation cannot be established if the audience cannot discover the speaker or writer's intention (Stokke 2010). The idea behind this is due to the conception of how an individual chooses signs. He first decides what he desires to refer to in his communication. He then chooses a sign that will – in his opinion, within the given context, and for the targeted audience – enable his audience to establish a reference relation between the sign he uses, and the object he intended the sign to refer (Bach 1994). Since he decides before using the sign and subsequently uses it, that is, he intentionally uses the sign. Clearly, this understanding of the role of intention in reference is not as crude as to say that reference cannot be formed without mind-reading. The debatable point is instead whether the audience *must* hypothesize the speaker's or writer's intentions to form the reference relation, and similarly, whether the intention or audience's conception thereof is part of context. Gauker's criteria make no apparent use of speaker's or writer's intentions. He avoids these mainly because the audience has no apparent way to access to what the speaker or writer have inside their mind. If an audience member claims to know the intention of the speaker, he is mistaken,



because he only knows some elements of the context, which he then mistakenly attributes to the speaker's or writer's intention. In broader terms, the important idea is that in the absence of mind-reading, any audience member is obliged to establish the reference relation only on the basis of whatever is accessible to him; the mind of another person clearly is not. Functional magnetic resonance imaging can reveal the parts of a human brain that tend to activate when meaning and truth of sentences are evaluated (Hagoort et al. 2004), but this is very far from the technologies that would allow something of a true intention of the speaker or writer to be known by another person. The takeaway here is that intention will not be considered as part of the context, and it will be taken as unknown when reference is established; whatever is known, and may seem as intention is merely the audience's conception of that intention that itself arises from elements of context, which must be accessible to the audience.

Limits to what the individual is aware of will be determined by his senses and knowledge, along with his limited ability to recall all these elements and articulate them in a meaningful way under the constraints (e.g., time) he obeys when forming the reference relation. They also work only under roughly speaking, normal conditions. For instance, they fail in case of some medical conditions. Oliver Sacks, a neurologist, describes in *The Man who Mistook his Wife for a Hat* the case of a musician who seems to need much more cues than the average person to make the reference relation:

It was while examining his reflexes... that the first bizarre experience occurred. I had taken off his left shoe and scratched the sole of his foot with a key — a frivolous-seeming but essential test of a reflex — and then, excusing myself to screw my ophthalmoscope together, left him to put on the shoe himself. To my surprise, a minute later, he had not done this.

'Can I help?' I asked.

'Help what? Help whom?'

'Help you put on your shoe.'

'Ach,' he said, 'I had forgotten the shoe,' adding, *sotto voce*, 'The shoe? The shoe?' He seemed baffled.

'Your shoe,' I repeated. 'Perhaps you'd put it on.'

He continued to look downwards, though not at the shoe, with an intense but misplaced concentration. Finally his gaze settled on his foot: 'That is my shoe, yes?'

Did I mis-hear? Did he mis-see?

'My eyes,' he explained, and put a hand to his foot. 'This is my shoe, no?'

'No, it is not. That is your foot. There is your shoe.'

'Ah! I thought that was my foot.'

Was he joking? Was he mad? Was he blind? If this was one of his 'strange mistakes', it was the strangest mistake I had ever come across... He also appeared to have decided that the examination was over and started to look around for his hat. He reached out his hand and took hold of his wife's head, tried to lift it off, to put it on. He had apparently mistaken his wife for a hat! His wife looked as if she was used to such things... We 'see' how things stand, in relation to one another and oneself. It was precisely this setting, this relating, that [the musician] lacked (though his judging, in all other spheres, was prompt and normal). (Sacks 1998, Part I, Chap. I)

The shoe is salient in the conversation between Sacks and the musician. Sacks mentioned the shoe before, so there is prior reference, and he points to the shoe, so



there is pointing. What seems missing are relevance and charity, while the location in the series criterion is irrelevant in this case, since the shoe is not part of a series. If you believe there is something to mind-reading, then the musician also seems unable to hypothesize Sacks' intentions. Of course, as Gauker himself observes, his criteria are not all there is to the formation of reference. They are interesting guidelines nevertheless, given the apparently slippery nature of the reference relation and the mechanism it operates on.

At this point, have we improved or worsened the understanding of context? Not unlike a fool who redefines by replacing one vague term with another, it can seem to some that we have moved from context as something, to context as something else, while we cannot tell much of the second something either. This would be an unfair evaluation, since we did start off from the reference relation as something between the two relata, sign and object, to a description of at least some conditions that appear required for that reference to form. Context is there an enabler of reference. For reference to form, the individual will be influenced by cues, which together form the context, while such cues include: (1) his own concepts (i.e., properties he predicates to objects); (2) his knowledge of how the sign of interest was previously used by himself and others; (3) salient objects; (4) previously mentioned objects; (5) objects relevant to the aims of the ongoing communication; (6) his background knowledge which will discard unreasonable sign-object pairs (this comes from Gauker's charity criterion); (7) objects pointed to; and (8) objects next in a series, if the communication involves a series of objects. With this, we have given conditions for something to be included in context. In other words, we told the constitution of context, or equivalently gave an intensional definition of members of context. A context are all such members, under usual conditions (i.e., in the absence of neurological and similar disorders that impede on the ability to use any of the criteria for reference).

A concept lumps together properties, whereby these properties take values: being a mathematician is a property that seems to take either a yes or a no as its value (i.e., it makes little sense to speak of quarter-mathematicians), while weight usually takes a positive real number as its value. An object is said to be an instance of a concept, and thereby stand in the instantiation relation to that concept if and only if the individual using the concept predicates to the object the qualities that correspond to the values of the properties of the concept. If we have a concept defined over a single property, weight, whereby the values of that property must strictly be over (i.e., not equal or less than) 1,000 Kg, then *any* object weighting strictly over a 1,000 Kg is an instance of the concept in question. The more properties we put in a concept, and the more restrictive values we impose on these properties, the more criteria we impose on an object to be an instance of the concept.

There are nuances with the instantiation relation (Gangemi et al. 2001). One subtlety involves the distinction between the instantiation relation and the set membership relation. The latter stands between a set and its member, while the former between a property and an object. If we say that *John is a mathematician*, did it say that John exhibits the property of being a mathematician, or that he belongs to the set of mathematicians, or both? There may be many cases where a property

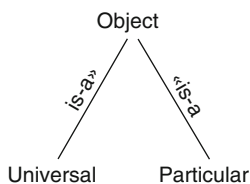
coincides with a set, as all members of a set may indeed share the same property. But this is not *always* the case, since we are not obliged to always build a set out of members who do share some same properties. Can we have a set of objects that are so different as to share no same property at all? One purist could argue that whatever we put in a set must share at least the property of being an object (in the sense we gave to the term *object* above). Another could say that all members of a set share at least the property of being members of that set. Both of these properties are, so to speak, trivial. The second complaint results from the set itself, so that the objects did not share it before being put in the set. More importantly, a set has no instances, while a property has instances. It is consequently said that a set can be an *extension* of a property, if that set contains all objects having the given property. But the object referred to by the sign *property* is not the object referred to by the sign *set*.

Another nuance lies in the distinction between instantiation and specialization. To say *this man is a living person* can be understood as saying that *this man* is an instance of the concept of *living person*, whereby that concept arises from the property of *being a living person*. Same with saying that *crow is a bird*. Put differently, the man has the properties of a living person, and the crow has the properties of a bird. The indexicals *this* and *that* are important in both examples, because reference is made neither to *any* or *some* man or crow, but specific ones. But then, when we say *all crows are birds*, does this mean that a property of being a crow is an instance of the property of being a bird? In *all crows are birds*, there are two properties, being a crow and being a bird. Although we have not seen all crows or all birds, we generalize nevertheless: if any crow is a bird, then the instances of *being a crow* are a subset of instances of *being a bird*. They are a subset, because there are birds other than crows. Here then, we have a subset relation: the set of all birds is larger than the set of all crows, so the latter is a subset of the former. In addition, we know that all members of the crows-set have the property of being a crow, while not all members of the birds-set have that property. There is consequently more than the subset relation between the crows-set and the birds-set. Namely, there is the information that the property of being a bird and of being a crow is true for all members of the crows-set. Hence, the extension of the property of being a crow refers to objects that are also members of the extension of being a bird. In moving from birds to crows, we increased the precision of our classification: while some properties are supposedly true for all birds, it is both these properties *and* additional ones are true for crows, but not penguins, pigeons, and so on. The extensions of the two concepts, *crow* and *bird*, stand in a subset relation, but the concepts themselves stand in the specialization relation: *crow* is a specialization of *bird*. As an aside, the extension of a concept is also called a *class*; the class of birds is thus the extension of the concept bird. To specialize, take a set, then identify a property that is not shared by all members of a set, but only some. That property serves as a tool to divide the initial set onto subsets. Doing so increased precision, since a more detailed description is given of the members of the initial set by the very act of identifying the property that divides. Remark that one specific crow is both an instance of the crow concept and of the bird concept, so the members of the extension of the crow concept certainly do not instantiate the members of the extension of the bird concept. Specialization is not instantiation.

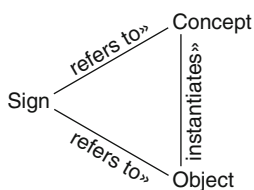
This brings us to yet another nuance, namely, about what can and what else cannot be instantiated. It is usual to call something that cannot be instantiated a *particular*. You cannot be instantiated, and neither can this specific ceramic vase nor that particular space shuttle. The concept of *living person*, *ceramic vase*, and *space shuttle* can be instantiated. If an object can be instantiated, it is called a *universal*. What is less clear at first sight is whether the property *being a crow* is an instance of the property *being a bird*. More generally, can an instance of a universal be a universal? If not, then any universal must have as instances only particulars, and we are confined to, so to speak, two levels only: particulars below, universals above (or the other way around, as you wish), and the instantiation relation between them to indicate that a particular is an instance of a universal. To see why the negative answer fails, consider the following line of reasoning. Suppose we could place all properties in a set. Then the object referred to by the sign *being human* is a member of that set. If then, any member of that set is a property, then that set is the extension of *being a property* property. One specific property, such as *being a human*, is then an instance and not a specialization of the object referred to by the sign *property*. We can thus have universals as instances of other universals.

Once the specialization relation is available, the sign–concept relation can be dealt in a fairly straightforward manner. We already said that a concept is an aggregate of properties; it lumps properties together, and for all practical purposes, it can be pictured as a list of properties. Consider again the object. We said that anything is an object, so that the term *object* is merely a generic sign for anything we may be communicating about. We can thus call any property by the term *object*, just as we can call a concept by that same term. Same applies for some specific sign – we can call that sign generically by the term *object*. Object in the sign/object/concept triangle is a catchall notion. It is consequently useful to supplement the sign/concept/object triangle by the taxonomy over these three objects. Each line in Fig. 3.3 is a specialization relation. Since any concept is a set of properties, and a property is a universal, we can have a property as the target of the reference relation, or a concept as that target. Consequently, the sign–object relation is the reference relation, in which the sign refers to a property or concept.

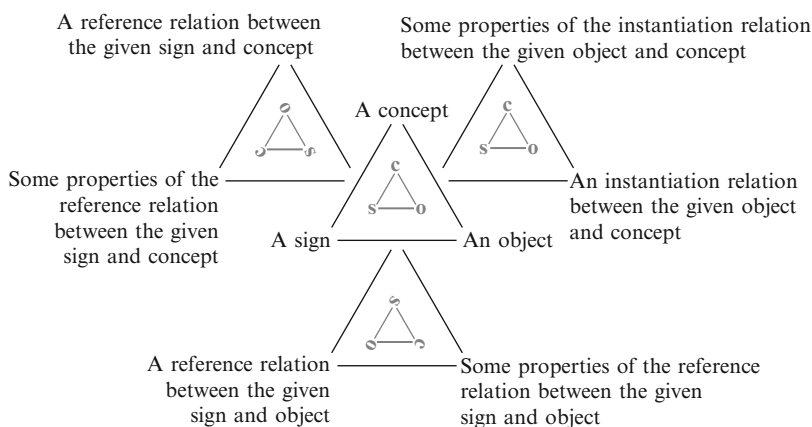
Figure 3.4 adds the names of the relations to the sign/concept/object triangle and tells how to read these relations. Figure 3.5 gives summary answers to the questions



**Fig. 3.3** Any object is either a particular or an universal. The specialization relation is denoted by the sign *is-a*. The criterion for specialization here, that is, the criterion for the distinction between universals and particulars is the ability to instantiate: a universal can be instantiated, while a particular cannot



**Fig. 3.4** Names of relations in the sign/concept/object triangle. *Double arrows* next to the names of relations indicate the direction in which to read the relation (e.g., *Sign refers to Concept*). A reference is both the sign–object and sign–concept relation, while the object is said to instantiate the concept



**Fig. 3.5** Figure 3.2 revisited. The central sign/concept/object triangle differs from the one in Fig. 3.2: the vertices here are some specific sign, some object, and some concept. The vertices of the outer triangles give answers to the questions in Fig. 3.2

asked in Fig. 3.2. It is necessary to speak, in Fig. 3.5 of *a* sign, *an* object, and *a* concept, instead of (any) sign, (any) object, and (any) concept (as in Fig. 3.1). We saw that reference forms in context between some sign and some object, or some sign and some concept. Precisely because context plays an important role, the properties of a reference relation are proper to that reference relation, that is, the reference relation between *that* given sign and *that* object (or that sign and that concept). If it is established in a given space and time that  $x$  refers to this object (e.g., which someone is pointing to), then that specific reference relation (i.e., the one between this object and the sign  $x$ ) has spatiotemporal properties that differ from the reference relation that could be established between  $x$  and this object at some other time and in some other location. Of course, when relata are different, we still have a reference relation, but one that has the property of relating some other sign and some other object. This gives a clearer reading of the phrase *Some properties of the instantiation relation between the given object and concept* in Fig. 3.5: these properties of the reference relation could be, e.g., the time and place at which the reference was established, or who formed that reference. Other such properties could be inspired

by the criteria for reference formation; e.g., a property of that reference could be its location in a series, or how salient the object was.

John Locke asked where at last to stop when we seek to clarify the understanding of a term. We look at its definition, then at the definition of those terms in its definition, and so recursively on, possibly without end. The triangle we managed to build here highlights this same problem, although in a somewhat different way. Given a sign, we can ask what it refers to: if we seek objects it may refer to, we are basically asking for an ostensive definition of that sign in the given context; if we are looking for the concept the sign is referring to, then we are looking for an intensional definition. The triangle and the white space it is drawn on, which hides the elements of context that influence the reference relations, both tell us that clarity cannot be the stopping criterion: a purist might start out by building the sign/concept/object triangle for *every* sign he encounters in the definition of another sign; his second step could then be to build the sign/concept/object triangle for every sign he encounters in each of the triangles he built in the first step; and so on. However enthusiastic this may seem, the stopping criterion Locke sought will remain elusive, because *any* sign in a natural language can only be defined by other signs in that same language. We can, of course, translate from one language to the next, but even several languages interconnected by translations suffer from that same problem that a single natural language has: namely, they have no truly primitive terms, which would require no definition at all, and thereby be understood equally by all.

### 3.2.4 *Sigs, Objects, and Concepts, in Advice*

What good is the sign/concept/object triangle in any analysis of advice? It seems that people make reference between signs and objects and concepts anyway, without knowing much of the troubles that have gone into elucidating the three corners of that triangle, and the relations connecting these corners. While we will consider other uses of the triangle in studying advice, it has an obvious role, which needs no definition of advice to be exemplified. Namely, once an individual knows that the complete set of parameters that determine reference is elusive, he should be careful both when dispensing and taking advice. A buyer of advice is expected to make reference between the signs used to communicate that advice, and some objects, so that the recommendation on the course of action does indeed play a role in the evaluation of alternative courses of action. It is relevant for the buyer/receiver to doubt any reference he makes, for different references need not be compatible with same actions. His doubt will be helpful to the extent he is able to ask questions about the sign, the reference relation, and the candidate objects and concepts, then seek answers. That may well be the way to avoid transforming one advice into another because of some unexpected reference: the giver/seller of advice refers to some objects and concepts, while the buyer/receiver refers to others.

Reference forms through context: the elements of context, that is, objects satisfying the various criteria discussed earlier are those that enter into some, currently

unknown equation that gives out the object that a sign of interest is referring to. Once an individual forms a reference relation, doubt of that relation will lead him to ask what other objects could have been at the other end of the reference, opposite of the sign. He could relevantly wonder whether some other objects, which also satisfy the accessibility, charity, prior use, and other criteria, could have been what signs in advice refer to; then ask how would this change of reference affect what he understands as the recommended course of action. The interpretation of advice will be discussed in considerably more detail in the fourth chapter, after we finally do give a definition of advice. It is nevertheless interesting to give a sneak preview of how very simple questions, arising from the understanding of the interplay of context and reference, can be relevant to the individual who listens to, and aims to understand advice.

As one of the prominent mainstream movies of the 1980s, Oliver Stone's 1987 white-collar crime drama *Wall Street* culminates perhaps not toward its end, but somewhere in the middle, in a scene that prominently features the term *greed*. The *Oxford English Dictionary* says that greed is an inordinate or insatiate longing, especially for wealth, which involves avaricious or covetous desire. Greed understood in such a day-to-day way is presumably one of the key topic of the movie, in which a young stock trader played by Charlie Sheen meets a big stock speculator, played by Michael Douglas. The speculator offers the broker to let him into the big deals on Wall Street, asking him to illegally seek insider information on such deals in return. The story is the usual one and includes no real surprises: it pits a seemingly naïve and well-intentioned individual to his opposite. We see Sheen's character enjoying financial success through his conspiracy with Douglas' character, then questioning this, repenting, and finally moving to a happy ending by compromising in the end the evil speculator and bringing him to justice. It requires no elaborate thinking to conclude, as some reviews have, that the highest point of the movie is a speech the speculator gives before a shareholder assembly of a paper company called Teldar Paper, in which he wishes to buy a number of shares sufficient to change its current management and lead the company wherever he intends to. What is interesting for us here is that Douglas' character in his address dispenses advice and uses the term *greed* repeatedly, while never trying to explain what he means by it. It could be argued that any viewer has an idea of what greed is, and he can consequently save words by avoiding an explanation in that scene. This claim stands, since any scene involving Douglas' character is effectively a case of greedy behavior; up to the speech scene, the movie builds an intensional definition of a greedy businessperson. But what happens when this scene is taken out of context, such as when viewed without seeing scenes the part of the movie that precedes it? In that case, no definition for the term *greed* is available from the movie itself, so that the viewer is left wondering what Douglas' character is trying to refer to anytime he says *greed*. The speech in question goes as follows:

I appreciate the chance you're giving me, Mr. Cromwell, as the single largest stockholder in Teldar, to speak...well ladies and gentlemen, we're not here to indulge in fantasies, but in political and economic reality. America has become a second rate power. Our trade deficit and fiscal deficit are at nightmare proportions. In the days of the 'free market' when

our country was a top industrial power, there was accountability to the shareholders. The Carnegies, the Mellons, the man who built this industrial empire, made sure of it because it was their money at stake. Today management has no stake in the company. Altogether these guys sitting up there own a total of less than 3 percent and where does Mr. Cromwell put his million dollar salary? Certainly not in Teldar stock, he owns less than 1 percent. You own Teldar Paper, the stockholders, and you are being royally screwed over by these bureaucrats with their steak lunches, golf and hunting trips, corporate jets, and golden parachutes! Teldar Paper has 33 different vice presidents each earning over \$200,000 a year. I spent two months analyzing what these guys did and I still can't figure it out.

One thing I do know is this paper company lost \$110 million last year, and I'd bet half of that is in the paperwork going back and forth between all the vice presidents. . . The new law of evolution in corporate America seems to be 'survival of the unfittest'. Well in my book, you either do it right or you get eliminated. In the last seven deals I've been in, there were 2.3 million stockholders that actually made a pretax profit of \$12 billion. I am not a destroyer of companies, I am a liberator of them. The point is, ladies and gentlemen, greed is good. Greed works, greed is right. Greed clarifies, cuts through, and captures the essence of the evolutionary spirit. Greed in all its forms, greed for life, money, love, knowledge, has marked the upward surge of mankind — and greed, mark my words — will save not only Teldar Paper but that other malfunctioning corporation called the USA. . . Thank you.

How are the shareholders to understand what Douglas' character is trying to refer to when he uses the term *greed*? Remember, they have not met him as closely as the viewer of the movie; they have been given no intensional definition via the scenes that precede his speech. As he finishes his speech, the scenario of the movie tells that they give Douglas' character "[m]uch applause as he sits. . . a standing ovation; shouts of approval." But consider what is said in the speech, that is, what are the elements of context from which the shareholders form the reference relation? He criticizes that the management has little ownership in the company, the hints at inefficiencies, then at his own successes at extracting financial gains from other deals, and ends by a celebration of greed. He thereby offers something called greed as a way out of the current management's failures. What he calls *greed* is his solution, even though he criticizes in the same speech the management for their inordinate or insatiate longing, especially for wealth, which involves avaricious or covetous desire, which they manifest by both a low ownership of a failed company and their high salaries and other benefits. We then see the shareholders cheering his speech, and it is precisely at this point that anyone who understands that reference forms in context that this very scene can only be fiction. What can the shareholders refer the term *greed* to after his speech? What does he mean by *greed* in that speech? Careful listening only of what he says — combined with the knowledge that reference forms in context — will force the conclusion that the term *greed* may well be empty; what could he be possibly referring to in the usual reading of the term *greed*, after starting his speech by a critique of the management's behavior? That very behavior he criticizes falls under the usual reading of the term *greed*. In a way, he first rejects greed, then embraces it. Perhaps when he uses *greed* he should better use *his greed*, while *their greed* would be what he refers to when speaking of the management's inefficiencies and high salaries. If cheering marks the shareholders' acceptance of his advice (to take up his proposal for managing by *greed*), then they have made reference between the term *greed* and some object that fails the criterion of being



charitable, while they have also failed to see that the behavior of the management he criticizes is *greed*. If they are taking advice, which includes empty terms, then what is it that they are accepting? It may be that by accepting nothing, they are effectively accepting anything.

We are left wondering why these shareholders are cheering to what he said? If we take the cheers as a sign, what objects/concepts does that sign refer to? They refer to, or mark an event, his victory, yet another in a line of triumphs that started from the outset of the movie. If the cheers are removed, it remains unclear who won the argument, the management or Douglas' character. Cheers render it clear, they are a key element of that scene, that is, of the context, in which the viewer of the movie grasps that Douglas' character has won yet again. The cheering in that scene is not unlike the laugh track (synonymous with canned laughter) in televised situation comedies. In the latter, laughter is forced on the viewer as another element of context of reference – regardless of the viewer (mis)understanding the joke without it – so as to facilitate or force the viewer to make reference between the scene and amusement. In *Wall Street*, the scene would fail in having a strong effect on the viewer without the cheering crowd, since listening only to what Douglas' character says leaves the impression that he has not properly thought out his ideas, and his speech ends up being a messy argument that few can take as something that can seriously evoke such a cheering response from an audience.<sup>6</sup>

As a final remark on *Wall Street*, it is interesting to note how different times come with different qualifications of Douglas' character. That and other scenes have been recalled countless times in popular culture, while the behaviors depicted therein

---

<sup>6</sup> There are discrepancies between Stanley Weiser and Oliver Stone's script for *Wall Street* and the theater version of the movie. One of these differences is that the speech on greed is considerably longer in the script than in the movie, and we have no speech from the character called Mr. Cromwell, whom Douglas' character confronts in front of the shareholders. The speech on greed is actually a response to what Cromwell says at the same shareholder assembly in the script, but not in the movie: "The original fundamental reason for Wall Street was to capitalize American business, underwrite new business, build companies, build America. The 'deal' has now succeeded goods and services as America's gross national product and in the process, we are undermining our foundation. This cancer is called 'greed'. Greed and speculation have replaced long-term investment. Corporations are being taken apart like erector sets, without any consideration of the public good. I strongly recommend you to see through Mr. Gekko's [character played by Douglas] shameless intention here to strip this company and severely penalize the stockholders. I strongly recommend you to reject his tender by voting for management's restructuring of the stock." We get from this more about what is meant by greed in that debate. Douglas' character also says less than is in the script, where he comes across even as an environment-conscious businessman, when he says "[Teldar Paper's] crown jewels are its trees, the rest is dross. Through wars, depressions, inflations and deterioration of paper money, trees have always kept their value, but Teldar is chopping them all down. Forests are perishable, forest rights are as important as human rights to this planet, and all the illusory Maginot lines, scorched earth tactics, proxy fights, poison pills, etc. that Mr. Cromwell is going to come up with to prevent people like me from buying Teldar Paper are doomed to fail because the bottom line, ladies and gentlemen, as you very well know, is the only way to stay strong is to create value, that's why you buy stock, to have it go up." The theater version of the movie actually strips so much of the argument in front of the shareholders that the reference of *greed* could only be unclear.



have both been hailed and criticized.<sup>7</sup> When a viewer sees this scene during a period of financial boom, Douglas' character and his traits are referred to a different behavior, or at least a behavior bearing different properties (e.g., a desirable behavior, and not an undesirable one) than when he views it in times of financial crisis. This changing reference of the movie – or, to use a plain phrase, its openness to interpretation – can itself be variously understood. It could be a strength of the movie, making it timeless as it somehow makes itself interesting in any period. It could also or otherwise indicate that it is very hard to pin down some signs, such as greed, and that Stone fails in describing greed in the same way that Plato fails in trying to define beauty, justice, and courage.

### 3.3 Ontology and Primitive Terms

Joch Locke's problem of where to stop in elucidating a sign (see Sects. 3.1–3.2) makes the reliance on primitive terms a necessity whenever a definition is sought, or whenever advice is dispensed. Primitive terms seem like a cover-up: we give a definition of a sign via other signs and take the latter as primitive; we have effectively defined something via something else, whereby the latter remains undefined. This is done with the hope that the reader or hearer has an understanding of primitive signs close to ours. There is undoubtedly something of that cover-up in definition, but we can still discuss primitive terms, if nothing then to indicate what understanding thereof is *not* appropriate, or to somehow get closer to what objects and concepts the primitive signs refer to. The aim of this section is to see how such discussions of primitive terms can be made rigorously, and how sets of primitive terms can be built, which are then used in the design of definitions.

Avoiding the cover-up in practice equates to make explicit as much as feasible the assumptions held when designing a definition. One structured and fairly rigorous way to do so is by discussing the ontological choices behind the definition. Such choices are typically called *ontological commitments*, which have a double role to play here. First, it is difficult to criticize and revise a proposal for anything, including this very proposal for the definition and analysis of advice, if a considerable part of it remains hidden, that is, if its hypotheses are not explicitly stated in as much clarity as feasible, even though it is apparent that some sort of perfect clarity will remain elusive. We cannot seriously claim to offer an analysis of advice, if we cannot provide an analysis of the basic concepts, on which the assessment of advice builds upon. Second, an explicit account of ontological commitments illustrates in itself the difficulty there is to come to an agreement, especially when much of what

---

<sup>7</sup> Michael Douglas, interviewed by *The New York Times* in 2007, said he would not mind if he never had “one more drunken Wall Street broker come up to me and say, ‘You’re the man!’” (Cieply May 5, 2007) The speech is celebrated in other movies (e.g., the similarly themed 2000 crime drama *Boiler room*), mentioned in press (Akst August 4, 2002; Anonymous May 16, 2002) and by prime ministers (Rudd Oct. 6, 2008).

usually remains implicit, is actually rendered explicit as clearly and precisely as feasible. Discussing the ontological commitments here thus serves to illustrate discussions that can be done for *any* term that is mentioned in advice or meta-advice. A discussion of ontological commitments thus continues the preoccupation with the precision of definitions and reference.

### 3.3.1 *Ontology in Philosophy*

“To study the most general features of reality and real objects” (Peirce 1931, 1935, 1958) is to deal in matters of ontology in philosophy, entailing questions of what exists, or what relationships there are between whatever we could admit or establish to exist. In asking what there is, or what exists, the plain question is what something we would call reality or world is made up from. In doing so, as Thomas Hofweber, a philosopher, conveniently summarizes, ontology “tries to establish what (kinds of) things there are, the other half tries to find out what the (general) properties of these things are and what (general) relations they have to each other” (Hofweber 2005, p. 256) A question, such as *Are there numbers?* is an ontological question. It might seem trivial, however, since anyone who has had any basic training in arithmetic will answer it in the affirmative. To understand *when* – i.e., under which conditions – such a question rejects the trivial answer can be illustrated by the arguments that Rudolf Carnap, a twentieth century philosopher, voiced when he argued that questions such as *Are there numbers?* or *Are there universals?* actually do have trivial affirmative answers. Namely, once an individual is taught arithmetic, whereby he learns the signs and manipulations that arithmetic involves, he can only answer that numbers do exist. This seems straightforward, since the objects that the signs and techniques of arithmetic manipulate are called number: in the terminology we used in the earlier sections, we can say that the sign *number* refers to any object that has the property of being subjected to the techniques of arithmetic, which is a hard way to say that any object that the techniques of arithmetic manipulate is called *number*.

Questions of whether something exists come with trivially affirmative answers *only if* they are asked within a given framework, which presupposes that these objects do in fact exist. In other words, objects called numbers can trivially be said to exist within arithmetic, since arithmetic does manipulate something, and we have a long-standing convention to refer to these objects by the sign *number*.

Ontology is a discipline of metaphysics, and Carnap argued that metaphysicians ask a rather different question when they deal with problems of ontology. Namely, when they ask *Are there numbers?*, they are in fact asking whether objects called *numbers* exist outside of the framework, in which their existence is trivially affirmed. Their question is not *Are there numbers inside arithmetic?* but rather something like *Are there numbers outside (i.e., independently of) arithmetic?* The question of existence in ontology is whether objects exist “in reality” or “in the world” as opposed to existing only within a framework, which incorporates some manipulation techniques. Carnap went on to argue that such ontological questions

make no sense, since it is only within some given framework that existence can be affirmed or rejected (Carnap 1956). His argument rests strongly on the distinction between questions of internal and external existence:

If someone wishes to speak in his language about a new kind of entities, he has to introduce a system of new ways of speaking, subject to new rules; we shall call this procedure the construction of a linguistic *framework* for the new entities in question. And now we must distinguish two kinds of questions of existence: first, questions of the existence of certain entities of the new kind *within the framework*; we call them *internal questions*; and second, questions concerning the existence or reality *of the system of entities as a whole*, called *external questions*. (Carnap 1956, Carnap's emphasis)

In light of the internal/external distinction, we are led to wonder whether it is possible at all to ask and answer external questions of existence, since we always seem to be inside some framework, a very general one being the natural language we use to communicate, and the intellectual or otherwise tradition we have been brought up in. This problem has been discussed repeatedly since Carnap and well before him by, e.g., Friedrich Nietzsche (Nietzsche 1986), although seldom in the same way (i.e., using the internal/external distinction). Carnap seems to have the upper hand then, and doing ontology seems a mistake. Problems of ontology could then be rejected as pseudo-problems, which consume, but give back nothing of use. It is important for the rest of this section to understand a nuance here, which is in what way Carnap argued against ontology in philosophy. The nuance has to do with the role that ontology plays, next to scientific method for the acquisition of knowledge about the world. Scientific method incorporates criteria for the evaluation of the claims it offers, such as their testing through experiment. Within the body of knowledge about the world, scientific method has no special status, in the sense of its hypotheses being spared from criticism and revision, as illustrated in, e.g., *The Logic of Scientific Discovery* from Popper. Scientific method has properties that are hard to put to test via the scientific method itself; e.g., putting one scientific method through comparative experimental evaluation with other scientific or other method indeed seems elusive. Another way arose to deal with such properties: “[a] traditional approach to this was to say that philosophy articulated knowledge claims in this area, using special non-empirical resources such as metaphysical intuition. . . Science articulated truths about reality, based on observation; philosophy articulated truths about the structure of knowledge presupposed by science.” (O’Grady 1999, pp. 1021–1022): It is said that Carnap rejected this picture of the interplay between scientific method and philosophy (and thereby ontology). Instead of having the role described above, philosophy offers, clarifies, and revises frameworks for conducting empirical science and organizing knowledge acquired thereby. As new observations arrive to invalidate prior conclusions, frameworks get revised or replaced; they have no definite status, but only temporary, until another framework proves more appropriate according some criteria (e.g., it can accommodate observations that were exceptions in a prior framework). This is a denial of a knowledge that is somehow superior to knowledge derived through scientific method. A framework is “an artificial model built to sharpen and highlight issues which are less clear in scientific inquiry.” (O’Grady 1999, p. 1025)

The very idea of rejecting the ambitious role for metaphysics and ontology in philosophy as giving foundations for scientific method lets us build frameworks as ontologies (often in such way that an ontology forms part of the framework), in ways we discuss in the rest of this section. Instead thus of asking, e.g., *Do numbers exist?*, an ontology containing something so named will be built only if the designer of that ontology has a use for that something, which he calls *numbers*. By building that ontology, its designer commits to the existence of objects called numbers inside his framework, since he needs these objects to achieve whatever aim he has set himself. Such an *engineered ontology* opposes very strongly the notion of an *absolute ontology*, which would describe reality as it, so to speak, truly is. The idea of the absolute ontology is the one that there exists a unique truthful description of the world, which is made of framework-independent particulars. This relativization of the role of ontology within metaphysics, and of the role of philosophy in relation to scientific method still begs the question of whether the universals we refer to in the ontologies we engineer (e.g., *numbers* in arithmetic, *beliefs* in epistemology, *emotions* in psychology) do have real, or in-the-world instances. Even if an individual rejects the role an absolute ontology would play for scientific method, he still has to answer how he believes his concepts to be anchored in what is typically called reality. Given a sign then, and a concept it refers in an engineered ontology, is the object in the sign/concept/object triangle only an illusion?

Any answer to the question of how the universals of an engineered ontology reflect reality – if they have real instances, or their instances are illusions, or some in-between – requires taking a position on how individuals learn about reality. Hilary Putnam, a philosopher argued that “elements of what we call ‘language’ or ‘mind’ *penetrate so deeply into what we call ‘reality’ that the very project of representing ourselves as being ‘mappers’ of something ‘language-independent’ is fatally compromised from the very start*” (Putnam 1990, p. 28; Putnam’s emphasis).

The language used for description already imposes some structure to reality, by influencing the classifications the individual makes and thereby how he distinguishes, refers to, and relates whatever it is that he has access to in his own context. Roughly put, the idea is that there is a filter – made of language and mind together – that mediates the individual’s conception of the very reality that surrounds him and that his conception is about. Such *conceptual relativity* that Putnam argues for “arises because our theorizing. . .depends on the concepts we use to think about the world, and these concepts ‘cut the world up’ in some particular way. The world itself, considered independently of these concepts, does not determine any one way of ‘cutting things up’, and different schemes of concepts [i.e., engineered ontologies] fill this role in different ways, producing various solutions for our highest level theoretical problems and the questions of ontology [in philosophy] that go along with them. . .Given ‘conceptual relativity’ there cannot be a single true description of a completely independent reality. . .because the world does not determine answers to basic ontological questions independently of our variable conceptual assumptions” (Anderson 1998, p. 1).

An individual’s descriptions are relative to his own engineered ontology, and it is hard to argue that such a framework is *not* contextual and thereby not relative

to the individual, as it is formed and changed over time and from his own *prima facie* evidence. *Prima facie* evidence for an individual is another name for any sign that refers to objects, whereby these objects are part of this individual's experiences and/or observations (Bealer 1992). Depending on what philosopher one sides with, experience could include only sensations, or only observations, or both, some would add reflections, such as introspection and intuition, and others also take in the experiencing of emotions.

To accept conceptual relativism leads us to say that the primitive terms, that is, the primitive signs – with which we shall construct the definition of advice later on – do not refer to some universals in an absolute ontology. Rather, they refer only to those universals to which we do commit explicitly (i.e., we write down that we do commit to them), and which we describe to a feasible extent (i.e., give their intensional or ostensive definition). Commitment to the universals in the engineered ontology can come only if we are ready to discuss arguments in favor and against each universal individually and all of them together, and only after such discussion is explicit, and left open to criticism. No scientism<sup>8</sup> is in this way allowed, as no part of an engineered ontology has some privileged status that protects it from revision. Neither objective facts, nor absolute universals, then are sought. Instead, the universals are to be offered with an attitude of a practical doubt, which not only seeks to reject any or all of these universals when countering evidence and argument are overwhelming, but when it does so, it commits to engineer new ontologies until some further countering argument or evidence comes along, requiring that the new replaces the old yet again.

The sign/concept/object triangle is thus not real in the absolute ontological sense, and neither will be the universals we will be introducing throughout the rest of the chapter. The triangle is itself an engineered ontology, which summarizes some *prima facie* evidence for the separation between signs, concepts, and objects. In that same sense, the reference relation, and all we said about its formation in context, is a summary of *prima facie* evidence, accumulated and filtered down through, usually rigorous, discussion. Do we not fail in saying this the very project of defining and subsequently analyzing something called *advice*? Have we not tore down the very presumption of relevance of such an endeavor by admitting to conceptual relativism and the conventional nature of the very conceptual tools we intend to use? To be sure, we did reject the notion that these engineered ontologies somehow connect to real objects, or, so to speak, that *reference* or *sign* exists within an absolute ontology. An engineered ontology is never an absolute ontology. What makes engineered ontologies relevant – beyond their role in packing together and relating the universals they refer to, while explaining what commitments are made and why – is that they often reflect some *prima facie* evidence that *recurrently* preoccupies. The concerns in such an engineered ontology do not ordinarily arise out of whimsical *prima*

---

<sup>8</sup> For *scientism*, the *Oxford English Dictionary* says that it is “[a] term applied (freq. in a derogatory manner) to a belief in the omnipotence of scientific knowledge and techniques; also to the view that the methods of study appropriate to physical science can replace those used in other fields such as philosophy and, esp., human behavior and the social sciences.”

*facie* evidence, but from something that seems to (approximately) fit *prima facie* evidence, experienced or observed by potentially many people, and which they find important enough to invest in a debate. In case of the reference relation, for example, we have seen that it has been, and remains a topic of discussion for considerable time and between many. What further must be emphasized is that conceptual relativism does not entail a strong skepticism, which would have it that nothing believed can be verified as true and prefer an idle mind to one that is skeptical but still interested in answers, imperfect as they may be, yet amenable to refinement. This echoes Putnam's remark on the status of truth, which opposes saying that true is only that, for which we have no evidence or argument against, but have some evidence or argument for, that is, true is what is justified, or rationally acceptable:

To reject the idea that there is a coherent 'external' perspective, a theory which is simply true 'in itself', apart from all possible observers, is not to *identify* truth with rational acceptability. Truth cannot simply be rational acceptability for one fundamental reason; truth is supposed to be a property of a statement that cannot be lost, whereas justification can be lost. The statement 'The earth is flat' was, very likely, rationally acceptable 3,000 years ago; but it is not rationally acceptable today. Yet it would be wrong to say that 'the earth is flat' was *true* 3,000 years ago; for that would mean that the earth has changed its shape. In fact, rational acceptability is both tensed and relative to a person. In addition, rational acceptability is a matter of degree; truth is sometimes spoken of as a matter of degree (e.g., we sometimes say, '*the earth is a sphere*' is *approximately true*); but the 'degree' here is the *accuracy* of the statement, and not its degree of acceptability or justification.

What this shows, in my opinion, is not that the externalist view is right after all, but that truth is an *idealization* of rational acceptability. We speak as if there were such things as epistemically ideal conditions, and we call a statement 'true' if it would be justified under such conditions. 'Epistemically ideal conditions', of course, are like 'frictionless planes': we cannot really attain epistemically ideal conditions, or even be absolutely certain that we have come sufficiently close to them. But frictionless planes cannot really be attained either, and yet talk of frictionless planes has 'cash value' because we can approximate them to a very high degree of approximation.

Perhaps it will seem that explaining truth in terms of justification under ideal conditions is explaining a clear notion in terms of a vague one. But 'true' is *not* so clear when we move away from such stock examples as 'Snow is white.' And in any case, I am not trying to give a formal *definition* of truth, but an informal elucidation of the notion.

The simile of frictionless planes aside, the two key ideas of the idealization theory of truth are (1) that truth is independent of justification here and now, but not independent of *all* justification. To claim a statement is true is to claim it could be justified. (2) truth is expected to be stable or 'convergent'; if both a statement and its negation could be 'justified', even if conditions were as ideal as one could hope to make them, there is no sense in thinking of the statement as *having* a truth value. (Putnam 1981, pp. 55–56; Putnam's emphasis)

We can thus take conceptual relativism seriously, without thinking that it is so relative that truth loses its role. Empirical evidence can have a role in moving from rationally acceptable beliefs to truth, and justification/rational acceptability (which we shall revisit later) have a role to play with regards to truth. Although an absolute ontology is thereby avoided, engineered ontologies are acceptable under conceptual relativism. We can move here readily to advice and remark that none of it

will be given within an absolute ontology, so that any advice or meta-advice will be meaningful to an individual only to the extent that he shares (some part of) a conceptual schema, an engineered ontology with those offering him the advice. Hence the importance of trying to engineer ontologies and elucidate the ontological commitments therein.

Conceptual relativism cannot work without postulating that people have different *conceptual schemas*. These, as Donald Davidson, a twentieth century philosopher, summarizes, are “ways of organizing experience; they are systems of categories that give form to the data of sensation; they are points of view from which individuals, cultures, or periods survey the passing scene” (Davidson 1973–1974, p. 5). In our terms, they are engineered ontologies, related universals, which arise out of the individual’s *prima facie* evidence, the confirmation or disconfirmation thereof from subsequent evidence, in both of which we can find influences proper to culture, education, period, and so on. Under conceptual relativism, a conceptual scheme, or engineered ontology, is what *prima facie* evidence gets through to be classified and related to prior experience, or broadly speaking, understood. We should be careful not to see a conceptual schema as too rigid, for it is supposed to change to accommodate evidence that may counter prior (parts of) such a schema. Not only do *prima facie* evidences move through the schema, they modify it as they pass. To take conceptual relativism seriously begs the question of how relative is it, or in other words, “how much” do conceptual schemas of individuals differ? To ask the question in this way suggests that we could, say, take two people and somehow compare their schemas, then describe the differences. Since we admitted earlier that mind-reading will not do, we could attempt to have individuals make their schemas explicit. This seems a relevant endeavor, but will hardly go as far as we would want it, for there is some knowledge that is not only tacit, but very hard – if impossible – to make explicit (Polanyi 1958). Parts of the schema will remain locked in, and a perfect match between schemas will remain elusive. Clearly, if some exceptional degree of relativism is allowed, we run into a practical problem, as communication becomes only an exchange of sounds with no meaning passing through. Since this counters what we quite obviously experience to be the case, namely that some agreement is possible and that people do make honest claims to understand each other. Of course, how deep agreement is can be indefinitely questioned. To do so, however, would be besides the point, since when joint actions is needed for, e.g., survival, then communication does allow coordination, and ultimately, conceptual schemas cannot be too different after all. As Davidson says, “we make maximum sense of the words and thoughts of others when we interpret in a way that optimizes agreement (this includes room...for explicable error, i.e., differences of opinion)” (Davidson 1973–1974, p. 19). There is a limit to how much schemas can diverge, and while we cannot put some value on it, we could – reasonably it seems – accept that others’ schemas may finally not be so much different than our own. With this in mind, we can move on to discuss how ontological engineering can proceed, and how – by performing it – a set of primitive terms can be obtained, for subsequent use in a definition of advice.



### 3.3.2 *On the Engineering of Ontologies*

There are various occasions in Lewis Carroll's *Alice's Adventures in Wonderland*, when Alice politely asks questions to the strange creatures that she encounters, to better understand the strange place she found herself in, and what it is that they are telling her. At some point, she is talking to the disappearing Cheshire Cat on a tree, who says that wherever Alice chooses to go, she will meet mad people:

'But I don't want to go among mad people,' Alice remarked.  
 'Oh, you can't help that,' said the Cat: 'we're all mad here. I'm mad. You're mad.'  
 'How do you know I'm mad?' said Alice.  
 'You must be,' said the Cat, 'or you wouldn't have come here.'  
 Alice didn't think that proved it at all; however, she went on 'And how do you know that you're mad?'  
 'To begin with,' said the Cat, 'a dog's not mad. You grant that?'  
 'I suppose so,' said Alice.  
 'Well, then,' the Cat went on, 'you see, a dog growls when it's angry, and wags its tail when it's pleased. Now I growl when I'm pleased, and wag my tail when I'm angry. Therefore I'm mad.'

Alice goes on to debate a problem of reference with the cat, while we remain puzzled as to whether Alice accepted she was, indeed mad as the Cat thought. Regardless of who is mad, their conversation aims some alignment of conceptual schemas; it is quite apparent that this is one of Alice's main preoccupations after falling through the rabbit hole, as we see her repeatedly debating what things are, and how what they seem to be refers to the words she and the strange creatures use. If we replaced Alice with someone far more concerned with agreeing, or pinpointing sources of disagreement with the Cheshire Cat, that individual could go one step further than Alice. He would engage in the business of defining universals and their relations, while finding signs to refer to both. He would engineer an ontology.

Having to clarify signs and reference relations, along with the objects that they are supposed to refer to is unavoidable. A conceptual schema is determined by its owner's conception of the properties of objects referred to by the signs he uses within that schema. A physicist designs formulas (or more broadly, models) that relate abstract objects (e.g., neutrinos, photons, and so on), and does so from some empirically grounded or theoretically thought-out understanding he has himself come to be convinced of. As many schemas are intended to be shared, if only because they are controversial, it is not only well mannered to do one's best in explaining what the signs in the framework refer to, but is crucial for any use and discussion thereof.

An engineered ontology should not be equated with (some part of) a conceptual schema. An ontology can only refer to (some part of) a conceptual schema, in the same sense a sign refers to an object. The ontology will make explicit the commitments that its authors have about what objects are relevant to their purpose, what properties these objects have, and what relationships they participate in. If accepted by others within a domain of inquiry, the engineered ontology will act as a foundation for the exchange of information in that domain. It will have the practical



roles of facilitating the comparison, integration, extension, or the suggestion of new conceptual schemas.

How does one start making an ontology? Consider an actual example of ontology-building from philosophy:

Entities bound both in space and in time can be called *events* or *non-continuants*. They are entities defined by their spatio-temporal extension. The entity whose boundaries are given in all four dimensions is an event. An event is an entity that exists, in its entirety, in the area defined by its spatio-temporal boundaries, and each part of this area contains a *part* of the whole event. . . Events are the *only* substances of this ontology. Only they can have genuine proper names and be the subjects of predication. A description of the world in the language of the first ontology is a description of events, their properties and relations. . . The concept of a thing, or a substance, as any chunk or chunks bound in space and time (and which, therefore has spatiotemporal parts and can be sliced both spacewise and timewise) can be accommodated by various systems of thought. (Zemach 1970, pp. 233–234)

The ontology above is made from a commitment to the dimensions of space and time. That which obtains a unique value across the three spatial and fourth temporal properties is taken as the primitive building block, out of which all else is composed. That building block is called *event* and is primitive in the sense that – in this ontology – it cannot be broken down to smaller objects, i.e., there is no commitment in the ontology to parts of events: anything is either an event, or made of two or more events. From a methodological standpoint, this ontology is engineered in two steps. First, some backdrop is selected. From it are, so to speak, extracted four properties. Three of these are called spatial properties, and the fourth is called a temporal property. We should recall here the sign/concept/object triangle: e.g., *temporal property* is a sign that the author of the ontology wishes to use to refer to what we usually conceive of as time (which would be the object in the triangle, to which the sign *temporal property* refers), and to which we attribute properties (e.g., that it “flows”, that it “cannot be stopped”, and so on). The second step is to say that anything, any entity (i.e., any *object* in the parlance we adopted in this chapter) has these four properties, to which values are always selected. Once events are introduced as primitives, they can be related. Two events can stand in different spatiotemporal relations: e.g., one can occur after the other in time, next to the other in space. More elaborate concepts can be built by putting events together along one or more of their four spatiotemporal properties. For example, this desk is an event that “moves” through time (i.e., the value of its temporal property changes as time passes), while occupying certain values of spatial properties. In other words, we can use this ontology to speak about events.

With only one primitive, the ontology of events looks conveniently clean and simple. It has four properties that are accessible to anyone who ever had even a remote interest in geometry and basic physics. There are, however, some questions that it keeps unanswered, and which can cause trouble when we use this ontology to speak of and relate events. Namely, what is the unit (as this term is understood in physics) of the values of spatial properties, and what is the unit associated with the temporal property? Are there hierarchies of such units (e.g., millimeter, centimeter, meter)? If yes, what are the rules for converting values at one level of that hierarchy to obtain those of the next? If not, what is the one unit of choice (and why)? These

questions may seem like splitting hairs: an answer is that the ontology of events takes this from its context of use – in some cases, minutes are good enough, while at others, milliseconds might matter. If minutes are the unit of choice, then no event can last less than a minute: any shorter event will be rounded-up to the next minute. This will not always work, of course. If the ontology of events is used to provide a definition of *event* in a criminal trial involving several criminal acts (e.g., bank robbery, shooting of one witness, then shooting of another), it cannot leave open the question of units: if we take an hour as the unit for the temporal property, then maybe this is not fine grained enough – someone may rightfully argue that the hour can contain many of what that person calls by the term *events*, and distinguishing between them could be important enough to warrant the use of another unit. Same problem applies to the spatial properties. These choices determine how many events we can distinguish in a given time interval and some delimited space.

Tempting as it may be to call this ontology of events deficient, any such judgment should be made with care. Evaluation over at least one of the two kinds of criteria could conclude either failure or success, or, if the evaluator is so inclined, to some vague judgment in between. One criterion is whether the ontology fails with regards to the aim it has been designed with. The other set of criteria concerns whether there are errors in its design, which are unrelated to its purpose: e.g., it may mistake the instantiation relation with the specialization relation.

The purpose of an ontology dictates its depth. If the aim of the designer of the event ontology was to use this ontology to talk exclusively of, say, natural disasters – floods, earthquakes, tsunamis, and so on – as events, then the above ontology can only be the first step toward that aim. The next step would be specialize the *event* concept, by adding properties that will allow us to distinguish kinds of events. How *deep* an ontology should be depends on how detailed the classification it aims for must be. If we not only hoped to distinguish floods from earthquakes, but also, separate among each of these classes the subclasses that differ in terms of, e.g., the continent on which a natural disaster occurred, we can introduce a property, *continent of occurrence*, which can take as values the name of any of the continents. If we adopt the convention of seven continents, we will have seven subclasses for every class of natural disasters. All natural disasters are then events; all natural disasters are then divided into subclasses of earthquakes, floods, and so on; and finally, each of these subclasses is further subdivided into earthquakes in Asia, earthquakes in Africa, etc. As we specialize – as we partition the original class of objects – we are adding information to the ontology: any specialization requires that new properties be added to the ontology to act as criteria that partition the initial class. Increased specialization deepens an ontology; cutting out specialization relations from an ontology makes it shallower.

The purpose of an ontology dictates its scope. Speaking of scope, we could say that, if the purpose of the ontology of events is to talk *only* of events, then its scope is appropriate. Observe the circularity: if it is indeed the purpose of an ontology to be used to talk only of that which it includes, then clearly, its scope will *always* be appropriate. Any discussion of scope requires that the initial commitment allows us to distinguish what will be in and what left out of the ontology. This is the very

question of scope: what stays in and what stays out of an ontology? Once this is known, a judgment can be made on whether the scope is good enough to a given purpose of the ontology. Evaluating the scope thus, rather inconveniently brings back the problems of ontology in metaphysics: to know what remains outside an ontology, it is necessary to know another ontology, which *includes* the ontology being designed. To know if the ontology of events has an appropriate scope, we must know what else there is besides what that ontology calls an *event*. This is to ask either what *event* is a specialization of, or what not a specialization of *event*. When the initial object of the ontology – the object, which is specialized by the deepening of the ontology – is as general as the *event* defined above, then asking these last two questions borders on a discussion of metaphysics: what is an event a specialization of, and what is not a specialization of the event concept? Events are typically neither physical objects, nor facts, nor properties (Casati and Varzi 2008). Events differ from physical objects in at least five respects: (1) events are said to occur, while physical objects exist, and it is rather unusual to say that events exist, and that physical objects occur; (2) physical objects seem to have clear spatial boundaries and unclear temporal boundaries (e.g., this desk has a clear position in space, while it is unclear how long it will exist), while events (e.g., a birth) have no clear spatial boundaries, but possibly clear temporal boundaries; (3) physical objects occupy space, and two objects do not occupy the exact same space, while events can occur at the same time; (4) objects can move in spacetime, while events are fixed in spacetime; and (5) objects keep their parts through time, while an event accumulates its parts over time.<sup>9</sup> While both events and physical objects do differ, they are both not atemporal. Facts are often described as atemporal: if the death of Caesar did occur in Rome in 44 BC, then that fact does not change if it is considered today, or at any other time after that event took place. Our final distinction is between events and properties: once events are considered as particulars, they are not universals (as in Fig. 3.3), while properties are universals. It is clear from a discussion of events *versus* physical objects that engineered ontologies get never too far from ontological issues in metaphysics: any engineered ontology will rely to some extent on a debate in metaphysics; even when such considerations are not immediately visible (e.g., in the construction of an ontology of plants), it

---

<sup>9</sup> This fifth difference is best described by D. Hugh Mellor, who suggested this way of separating physical objects (he calls them *things*) from events: “Examples of temporally extended events are: wars, meals, avalanches, eclipses, explosions, nuclear reactions and the expansion of the Universe. All these events take time, and none is wholly present at any one time. An instant of time indeed contains no part of such events; it merely separates temporal parts of them, as an internal surface separates the spatial parts of a thing. It takes a stretch of time to contain any temporal part of an event. An event can be said to be present at a time only in the sense in which I am present at every point of space within my body. Things, on the other hand, are wholly present at every time at which they exist at all. Examples of things are: nations, people, omlettes, mountains, planets, bombs, atoms and galaxies. None of these things has temporal parts in the way it has spatial parts. It is, for example, nonsense to say that only a youngish part of Sir Edmund Hillary climbed a rather older part of Everest. Hillary himself climbed Everest itself, and both parties were wholly present during all temporal parts of that historic event.” (Mellor 1980, p. 283)

is a matter of omission, not of irrelevance. Does this mean that Carnap's critique of existence-outside-of-a-framework must be rejected? No; the sole consequence is that discussions of ontologies in metaphysics cannot summarily be rejected: they can give arguments for abstract distinctions (e.g., event *versus* physical object), even when the commitment to such distinctions is merely one of convention, and not the commitment that these distinctions, so to speak, truly exist in the world (i.e., outside of the engineered ontologies, in which they are put to use).

As soon as efficiency becomes a criterion in any process of ontology engineering – as some conception of efficiency almost always is – it is impractical to discuss in exceptional detail what is left *outside* an engineered ontology, e.g., of plants. Doing so would again pose the questions of why a plant is not an event, or a fact. Many questions would need to be answered in addition to those that concern the very classification of plants. Suppose for the sake of argument here, that *efficiency* is the ratio of the output over the input in a process; also, the input and output are assumed comparable, for their ratio would otherwise make little sense. No elaborate calculation is needed to see that less effort is needed to produce an ontology of plants if *only* questions about the concept *plant* and its specializations are to be answered, than if in addition, questions must be answered about what remains outside that ontology. This is the very strong reason why a distinction is made between *foundational* and *domain* ontologies. To design an ontology of plants is to design a domain ontology, that is, an ontology tailored to the domain of plant classification. A domain is not something that can be clearly and definitely delimited. Rather, its purpose is to establish, usually vague borders on the scope of an ontology. Sometimes it seems sharp enough to warrant little debate, as in the case of plants: *prima facie* evidence in many cases lets us easily distinguish, under usual conditions, plants from animals, or minerals, or bacteria. To make matters more difficult, there is an interplay between the domain and the ontology thereof. The very construction of an ontology affects the domain, to which the ontology is said to apply. We can start from the vague notion of plants, introduce properties into the ontology, define specializations, and so on. It is when borderline cases are considered – i.e., those objects that admit both arguments for and those against their classification as plants – that the borders of a domain are in question; the choices made at these times shape the limits of the domain together with the ontology. The distinction between a foundational and a domain ontology is not a matter of depth, but of scope. An ontology of plant is not a foundational ontology, because *prima facie* evidence tells us that there are objects that are not specializations of plants. As soon as *prima facie* evidence or perhaps a successful elaborate argument tell us that there are concepts, which are *not* specializations of the initial concept of an ontology, then that ontology is not foundational. If we can argue that there is something that is not a specialization of *object*, as this term is understood in this text, then the ontology in Fig. 3.3 is not a foundational ontology. Given the very wide scope of a foundational ontology, such an ontology will be built through ontological commitments on abstract distinctions, such as that between physical object and event, or event and property. The role of foundational ontologies is then clear: they will answer the questions about what remains outside domain ontologies. A domain ontology is then freed

from discussing again basic distinctions that it is not immediately concerned with. However, the engineer of a domain ontology will only be freed from such discussions if he classifies the initial concept of his domain ontology within a foundational ontology: when designing an ontology of plants, classify the plant concept within a foundational ontology. The flipside of the coin is that doing so commits the engineer to the ontological choices made in the design of the foundational ontology, so that not any foundational ontology fits with any domain ontology.

Choices of scope, depth, and foundational ontology are examples of *ontological commitments* (also called *ontological choices*); any decision made in the design of an ontology is an ontological commitment. Checklists of commitments to make would certainly be useful, framed as questions to answer in a step by step manner toward a new ontology. There are no standard checklists of such choices, just as there is not agreed upon method for the design of an ontology. The design of the ontology of events starts with commitments to three dimensions of space, and a dimension for time. There is then a choice to lump these four properties together into the concept of *event*. Further on, it is said that anything is either an event or a combination of events, so that physical objects are seen in a rather strange way, while properties and facts must remain outside. We see there a sequence of ontological choices that appear fairly clear, though also counterintuitive to some. It would be naïve to conclude from the description of an ontology the way in which it was designed. It is recognized that ontology engineering, or more fittingly – given how much of an art it is, rather than science – ontology design is a messy process. At best, it is iterative, involving repeated revision and clarification of its designer's initial ideas of what should be in the ontology, how these components relate to each other, and what their definitions need be. Consider the problem of delimiting the scope. Far from being a simple question to answer at the outset of some ontology design process, scope forms over the course of that very design process. Boundaries for scope are initially vaguely set, and it is through the subsequent ontological commitments on properties and specialization relations that these boundaries become sharp.

If there is no proven method for ontology design, what guidance is there? Without fixing the steps to take, and without saying exactly all the criteria that should be obeyed when making ontological commitments, it is possible to provide some criteria that are independent of the domain and foundational ontologies, and of the intended scope and depth of the future ontology. Involved in the design of ontologies and the design of methods for the design of ontologies, Nicola Guarino and Christopher Welty have suggested criteria for the evaluation of ontological decisions (e.g., Guarino and Welty 2002). The aim of such evaluation is to find out if an ontological commitment (which we entertain to take during design, or which may have already been taken) produces problems in the ontology. Problems can manifest themselves in various ways. One case is when we wish to classify some particular within a given ontology, and we have doubts as to whether it is an instance of a universal *X* or another universal *Y* in that ontology. If the individual doing that classification is not wrong himself, then that ontology fails to provide sharp criteria for distinguishing instances of *X* from instances of *Y*. Common rough ontologies

in economics and politics are the dichotomies, taken up often in media discourse, that oppose first to third world countries, or else, west to east, or north to south economies, or even developed and developing countries. Where are second world countries, if there are first and third? It is confusing to talk of Australia as of a north economy, which is clearly to the south of most north economies. Japan is then a western economy, even though it is to the east of most western economies. How sharp is the divide between developing and developed economies? The ontology that distinguishes developing from developed countries, or somewhat in parallel, north/south, and first/third world countries certainly subsumes that a sharp distinction exists. We should ask if there are meaningful ontological commitments behind such dichotomies. To be fair, the World Bank does supply a level of per capita gross national product as a line to separate developing from developed countries. It is unclear whether this same criterion applies to other of the mentioned dichotomies. At best, such ontologies are harmless; they are more likely to be harmful, however, as they certainly do stereotype and perpetuate prejudice (e.g., Hoeschele 2002). In any case, such rough ontologies require heroic ignorance from their users, or at least exceptional suspension of common sense.

Criticism of engineered ontologies should be carefully made: it is often very difficult to find properties and their values with which all ambiguity of classification can be avoided altogether, that is, which would allow any individual of interest to be classified without doubt in some class of the ontology.

An illustrative case is the design of a natural taxonomy, which aims to classify all known living organisms and thereby help name them. Before the 1970s, natural taxonomy relied on fossil records and detailed morphologies – i.e., properties that have as instances some elements of external appearance, such as shape, color, and pattern – to classify living organisms. This meant that in most cases similarities in external appearance led to the creation of concepts in natural taxonomy. Bacteria could not be classified in this way, as their morphologies are too simple to allow fine distinctions between very different microbes. As genome sequencing became feasible, the classification based on morphology was overturned as misleading, since it grouped together in some classes organisms that have evolved from different ancestors, and have sufficiently different genomes that such groupings do not reflect what some have called natural relationships among organisms. The usual animal/plant distinction became only secondary, while the first specialization is now onto bacteria, archaea, and eukaryota (Woese et al. 1990).<sup>10</sup> This case illustrates how moving from one set of properties (e.g., morphological ones) to another (genetic properties) changes fundamentally the classification of organisms, so that we are effectively dealing with two distinct ontologies, as the distinguishing criteria have changed. While a morphology-based ontology can still be useful for some purpose, it can no longer claim that its classes put together organisms that have followed

---

<sup>10</sup> *Eukaryota* refers to organisms with cells containing parts enclosed in membranes (e.g., the nucleus, which contains genetic material). *Archaea* refers to single-cell organisms that are genetically different from bacteria, although some morphological similarities exist. Genetic material is not enclosed in a membrane in the organisms classified under either bacteria or primitive bacteria.

similar evolutionary paths. The advancement of molecular biology and the technology it uses allowed new properties to be considered, and these in turn led to a new ontology. The previously designed ontology could not be reconciled with the new knowledge and was replaced with a newly designed one.

Change from one set of properties to another in the design of natural taxonomies also shows how important it is not to read too much from the specialization relation in any ontology. In a morphology-based ontology, the specialization relation between a general class, e.g., animal, and a special class, e.g., elephant, is there only to indicate (1) that all particular elephants share some morphological properties with other particulars of the animal class, and (2) that they do not share some other morphological properties with animals that are not elephants. As it says *only* that, it states neither that all particulars in the animal class have followed the same evolutionary path, nor that they all share some anatomical properties, nor that they may reproduce in the same way, and so on. It is only if we assume that morphological similarities are a sign of similar evolutionary origins that we can claim that natural taxonomy based on morphology indeed also reflects what is commonly called a tree of life. To make this hypothesis is to make the specialization relation refer to more than it actually does: it initially refers to the relation that stands between universals sharing some morphological properties; the hypothesis adds that these same relations also indicate that the *relata* share common evolutionary ancestors. Could we have argued that the specialization relation is overloaded (i.e., we make it refer to more than it seems to refer in the first place) in morphology-based natural taxonomies before the advent of economical genome sequencing? Of course, it is only with the advent of economical genome sequencing that the specialization relation in morphology-based natural taxonomies can be called overloaded with certainty. What cheap genome sequencing did not change, however, is that the specialization relation can only refer to the inheritance of morphological properties from the general to the special class (i.e., the elephant concept inherits the properties that define the animal concept), and that adding anything else are additional hypotheses. While *prima facie* evidence (here, simply looking at the morphology of different animals) indicates similarities and differences, it says nothing on whether they share ancestors in the evolutionary process.

An ontology classifies via universals, it is a collection of universals. Each universal in turn equates with a set of properties, so that we speak of ontologies having concepts as a synonym of ontologies having universals. If we wish to classify a particular in a given ontology, we must recognize its properties and see if they match to those of a concept in the ontology. Of course, we can only do this once the ontology has been engineered. In contrast, when we engineer an ontology, we need to choose the concepts and thereby properties via which we shall distinguish the particulars that we wish to distinguish. It is when we choose properties that we can ask two questions.

*Is a property  $p$  essential to the class of particulars?* If it is, then any particular of that class must exhibit that property to be a member of that class. If we say that *being white* is an essential property of swans, then any black swan will not be classified as a swan. This is clear enough for swans, since it is known that black



swans do exist: we consequently know that *being white* is not an essential property of swans, so that we should look elsewhere for their essential properties. Essential properties are of particular interest in definitions, since an intensional definition will be very interesting if it does rely on essential properties, and not any property (e.g., accidental ones, such as *being a one-legged swan*). Clearly, a special class cannot have essential properties that are in conflict with the essential properties of its general class.

*What properties are the identity criteria for a class?* Identity criteria are all essential properties that are sufficient to distinguish any member of that class from any member of any other class in a given ontology. For example, *being able to breathe* is an essential property of both humans and elephants, but is not an identity criterion if we only wish to distinguish humans from elephants for they both share it: *having two legs* is an identity property for humans, which is not satisfied for elephants. In natural taxonomy, having sufficiently common genome is an identity criterion for organisms, and it is on this basis that the ontology is designed. One way to discard lousy identity criteria is to consider plausible scenarios in which they may fail over time; if a particular can lose some property over time, then that property is not an identity criterion: *having two legs* is in this perspective not an identity criterion, for any human with one or no legs does not exhibit that property.

Design of ontologies is a messy, usually iterative process, in which a series of ontological commitments is made. Each of these ontological choices affects the scope and depth of an ontology, which are only known after the ontology is designed, when the designer of the ontology is content with his work. When a foundational ontology is sought as a root for a domain ontology, ontological commitments of the domain ontology must match those of the foundational ontology. Hence, the relevance of openly discussing ontological commitments. When properties are considered for inclusion in an ontology, we should determine whether they are essential, and if so, whether they are identity properties. This would ideally lead to an ontology that carries only identity properties and nothing else, for any other property is not necessary to make the distinctions that the ontology is intended for.

## 3.4 Advice, Defined

### 3.4.1 Initial Ontological Commitments

To act as a tool of coordination, advice must be transferred between individuals. The transfer happens via communication. It follows that only that which has been communicated or is being communicated can be called advice. Having been transferred via communication is thus an essential property of anything we might call *advice*. There is no advice without communication.

While advice is revealed via communication, this alone is not a sufficient identity criterion, merely a necessary one. If it were sufficient, then anything communicated



could be called advice. That would mean that absolutely *anything* an individual hears, sees, or experiences otherwise – i.e., receives through verbal or nonverbal communication – is a piece of advice. However, it is rather clear that a person's memory is not unlimited, and that, regardless of its limits, not anything that is stored in it can be recalled at any time and with same precision. In order to be called advice, some communication will also need to establish reference to some element of the decision situation, which is preoccupying the individual receiving that communication. This is the question of relevance: while it may be interesting to hear, say, that sharks attack surfers because they believe the latter to be turtles (apparently due to the visual similarity of the surfer sitting on a surfboard, when looked from below), this refers to no element of a decision problem, in which an individual has to choose an airline for his skiing trip; the former is simply irrelevant to the latter. Whether some communication is relevant depends on the context of the individual who receives advice: if he can establish a reference relation between the signs he hears, reads, or experiences otherwise, and the objects he perceives as accessible within his context, then what he is communicated may well be advice.

Advice must thus be a communication that is relevant. By relevance is meant that the individual must be able to establish the reference relation between the objects that are accessible to him and the signs used in that communication. Relevance is a subtle point here, since meeting properly the relevance criterion is particularly difficult and is likely to fail often to various extents. Consider political decision making in parliamentary systems, and in particular the debates in the US Congress on the matters of carbon dioxide emissions. Although the USA signed in 1992 the *United Nations Framework Convention on Climate Change*, which commits “to protect the climate system for present and future generations,” the US Congress had to subsequently pass any mandatory regulation on the emission of greenhouse gases. Among the regulations the USA ignored at the outset of the twenty-first century is the Kyoto protocol on emission targets for industrialized nations (PEW Center on Global Climate Change 2004). Many arguments have been voiced in the US Congress on the climate change debate and the adoption of emission caps. On 28 July 2003, a senator, chairman of the then-Environment and Public Works Committee, said the following:

Thus far, no one has seriously demonstrated any scientific proof that increased global temperatures would lead to the catastrophic predictions by alarmists. In fact, it appears just the opposite is true, that increases in global temperature have beneficial effect on how we live our lives. (United States Congress 2008, S10013)... What gets obscured in the global warming debate is the fact that carbon dioxide is not a pollutant. It is necessary for life. Numerous studies have shown that global warming can actually be beneficial to mankind.” (United States Congress 2008, S10019)... it would be beneficial to our environment and the economy. (United States Congress, 2008, S10022))

It certainly seems that this senator's remark speaks about carbon dioxide emissions, and that he has not missed the overall topic of the debate. It is, however, much more difficult to answer whether and how his communication is relevant to the hearers, that is, the Congress and Senate members in the sense of relevance that we have explained above. What is it that those *not* knowledgeable in carbon dioxide

emissions and the workings of global climate will refer these signs to? Clearly, the senator is a generalist with regards to that matter, as are the members of his audience. They are specialists in politics, not recent climate research. The references they will form will undoubtedly be questionable, as may their ensuing decisions. The role of his speech certainly is to advise, namely on how his audience should vote on the US legislation on carbon emissions. Beyond the very interpretation they may have, that is, the reference relations they will establish, and which we will revisit in the fourth chapter, the point here is to see that relevance asks us to look at *what* is communicated to study the advice that is given.

What we can conclude from the communication and relevance criteria is that any analysis of advice must involve an analysis of the communication that reveals that same advice. This observation very strongly influences the scope of an ontology of advice: instead of pulling out the concepts for an ontology of advice from thin air – while arguing, say, that precisely such and such concepts are useful or convenient – we can ground our ontology of advice in an ontology of content and form of communication. To ask what advice is, and what kinds of advice there are, thus requires some understanding of how communication happens.

If we thus have some idea of how communication may be happening, then we should be able to distinguish communication that reveals advice, from that, which involves no advice at all. A description of communication processes will thus provide the primitive terms for use in defining advice and the concepts that specialize it, allowing us further to accommodate the notion of relevance indicated above into a whole.

### 3.4.2 *Advice in Communication, Communication as Action*

Utterances that individuals exchange through communication amount to actions that advance their personal desires, intentions, beliefs, and attitudes. Content of communication can be conveyed in different ways by the speaker, entailing different effect on the hearer. This is reflected in linguistics by the separation, in any given statement, of the *dictum*, or what is said, from the *modus*, or how it is said in terms of speaker's attitude about what is said (i.e., the way it is said). This distinction finds early support in, for instance, Frege's *Begriffsschrift* (Frege 1879), where a distinction is made between the content of a sentence that expresses a judgment, and the act of judgment, that is, the assertion of the content. In the *Tractatus*, Wittgenstein also makes a distinction along the similar lines by stating that "...A proposition shows how things stand if it is true. And it says that they do so stand." (Wittgenstein 2001, paragraph 4.022). Stenius (1967) elaborated on Wittgenstein's initial suggestion by distinguishing between a "sentence radical" – i.e., the content – from the "modal element" that amounts to the mood of the sentence.

The dictum/modus separation is not uncontroversial. While Wittgenstein argued for it in the *Tractatus*, he changed his mind later in *Philosophical investigations*, where he doubted that content can be sharply split from the purpose of a

statement (Wittgenstein 1953). This change has been very relevant in twentieth century philosophy, and can be explained through a critique of an idealistic theory of language, and the passage from viewing language as something that is clean and well structured to seeing it purely as an instrument for doing things, a hammer of sorts. The idealistic vision of language says that every word has a meaning, and that this meaning is the object to which the word refers. Moreover, such meanings (i.e., objects referenced by words) exist independently of particular natural, mathematical, visual, or other languages, contexts, or whatever else could make meaning and reference malleable. To put it bluntly, there is something called language that is somehow independent of how the speaker uses it, and reference is definite: each word refers to specific objects, and there is nothing more to it. That this is nonsense is easy to recognize from *prima facie* evidence that people can manage to communicate while agreeing on changing the references of signs, as in, e.g., ciphers, and from the overall difficulty of reaching agreement on the meanings of signs. Now, what many philosophers attempt to do is build precisely such languages, which take the form of mathematical formalisms, and usually one logic or another. Many logics feature the mechanism of propositionally, which serves to relate sentences together via connectives, such as that of conjunction, or negation, or some combination of these two, and others. Not only is there a problem of relating words into sentences, but that of what meaning two sentences in conjunction have? Within a logic, this seems simple enough, since the precise rules (e.g., grammar, axioms, and/or a model theory) are known, but nothing of that kind of precision seems to apply in natural language. That a natural language is not a logic, and that it may lack a structure that even elaborate logics may have can only be supported when one asks what happens when words have no fixed meaning, that is, no fixed object of reference? As Wittgenstein puts it:

We see that what we call 'sentence', 'language' has not the formal unity that I imagined, but is this family of structures more or less related to one another.—But what becomes of logic now? Its rigor seems to be giving way here.—But in that case doesn't logic altogether disappear?—For how can it lose its rigor? Of course not by our bargaining any of its rigor of it. . . We are talking about the spatial and temporal phenomenon of language, not about some non-spatial, non-temporal phantasm. . . But we talk about it as we do about the pieces in chess when we are stating the rules of the game, not describing their physical properties. The question 'What is a word really?' is analogous to 'What is a piece in chess?' (Wittgenstein 1953, paragraph 108; Wittgenstein's emphasis)

If a word (sign in our sign/concept/object triangle) is an instrument, then it obtains reference through use, or more precisely, the context of that use. It is by knowing the rules of chess that we know, so to speak, the meaning of a piece in that game. In itself, the word alone has no meaning. It is a sign without reference and obtains the reference through use. This contradicts none of what we said earlier about the sign/concept/object triangle and conceptual relativism: contextual cues influence the formation of the reference relation over the course of communication, including the individual's experience of past uses, although without believing a unique initial baptism of an object by a sign.

The distinction dictum/modus thus seems to counter the idea that reference forms through context. The distinction divides the context into two parts, calling one the

dictum, and the other the modus. The usual examples of that separation are straightforward and seem not to cause too much trouble. When an individual says that a door should be opened, we may understand the contextual cues of his communication so that we attribute to that individual a desire that some particular door (to which he may be pointing) be opened; in the presence of other elements of context, we may see him as saying that some other individual should open the door in question, i.e., seeing him as issuing a command. It could be said then that there is here one dictum, and two modi: in the first, the individual conveys a desire, while in the second, he commands. Of course, any claim to a clear separation between dictum and modus involves at least some wishful thinking, a desire for a clean and sharp divide that would lend itself to a mathematical formalization, perhaps via some elegant mathematical logic. It is further evident that we need not reject the dictum/modus distinction summarily, for doing so would resemble to throwing out all of ontology when only ontology in metaphysics is rejected. Instead, we can hypothesize this separation as a convenient way to speak about what words and sentences may describe and the way that people communicate these descriptions.

Why do we need here a humbled distinction between dictum and modus? An individual's advice to another may be conveyed in different ways, to different aims: the advice-giver may wish to influence the recipient by suggesting new desires or changes to his existing ones, or aiming to change what the other believes is rationally justified, or further may wish to influence the emotions or moods the other has about alternative courses of action. As the individual receives advice, he can distinguish that which aims to influence his beliefs, from those that aims the same for his desires, or intentions, or attitudes. We can thus distinguish kinds of advice depending on these aims. The kinds of aims can be obtained from the kinds of modi we can identify in communication, and it is only for this that we keep distinguishing dictum from modus. In doing so, we will not admit that there exist (in the sense of metaphysical ontology) dicta and modi, merely that they can help in our other ontological commitments on a classification of advice.

What modi are there in communication then? One way to answer this is to see communication as a series of actions performed by the speaker and the hearer, as John L. Austin, a philosopher, argued in *How To Do Things With Words* (Austin 1962). The next step is then to find a classification of such actions. John Searle, another philosopher, suggested that a meaningful utterance in communication amounts to an attempt by the speaker to perform a *speech act* to convey something to the hearer. Communication is considered as action (Searle 1969): a speaker makes an utterance in an attempt to change the state of the world. What distinguishes speech acts from nonspeech actions is the domain of the speech act, that is, the part of the world that the speaker wishes to modify. With speech acts, the aim is to influence the mental state of the hearer. Elementary speech acts are of the form  $M(C)$ , where  $C$  is the dictum and  $M()$  is the modus in which  $C$  is communicated. Depending on modi, Searle distinguishes assertive, directive, commissive, expressive, declarative, and representative declarative speech acts (Searle 1975).

- *Assertive*. An assertive speech act indicates to the hearer that the speaker believes that the conditions (equivalently, states of affairs) described in the content of

the speech act hold. For example, the following quote of the same senator we mentioned earlier is an assertive: “Mr. President, the fact is this treaty [i.e., Kyoto protocol] is not based on sound science. The scientific community has not definitely even close to definitely – concluded that there is global warming caused by human actions. The science is inconclusive and often contradictory” (United States Congress 2008, S10309). The content of the assertion above can be summarized as the lack of definite scientific conclusion about the relationship between human actions and the phenomenon of global warming. By asserting this content, the speaker points out belief in its truthfulness. It is, of course, unnecessary for the content to be actually true; what the assertive conveys is only that it is deemed true by the speaker.

- *Directive*. The content of the directive speech act describes conditions that the speaker desires to see become true. In contrast to the assertion, the speaker believes that the conditions do not hold and desires that they become so at some potentially undetermined point in the future. For instance, corporate vision or mission statements typically amount to written-down directive speech acts, as the following one: “Enron’s vision is to become the world’s leading energy company – creating innovative and efficient energy solutions for growing economies and a better environment worldwide” (Anonymous 2000b, p. 4).
- *Commissive*. The commissive speech act indicates to the hearer that the speaker intends to perform actions described in the content; if the content describes conditions, then the commissive indicates that the speaker will perform actions needed to bring about states of affairs in which these conditions hold. Consider the following statement: “We are dedicated to conducting business according to all applicable local and international laws and regulations, including, but not limited to, the U.S. Foreign Corrupt Practices Act, and with the highest professional and ethical standards.” (Anonymous 2000b, p. 5) The intention is given by announcing the dedication to follow unidentified courses of action that have the same characteristic, that is, maintain some conditions true: namely, the respect of legislation applicable to the lines of business and geographic regions of interest.
- *Expressive*. An expressive speech act conveys the speaker’s attitude about a condition that may hold. For instance, when another senator said: “Our [i.e., US] military has done everything they have been asked to do, and they have performed excellently” (Pelosi February 13, 2007), this conveyed an attitude about the actions of the US troops in an armed conflict, here Iraq at the outset of the twenty-first century. Looking at the notion of attitude in psychology (Bizer et al. 2003) (which is what expressive speech acts convey), attitude is identified most closely with affect, i.e., a general evaluative reaction (e.g., “I like  $x$ ” or “I like  $x$  more than  $y$ ”). An attitude amounts to a description of an evaluation in terms of degree of favor or disfavor (Eagly and Chaiken 1996). Such degrees vary in sign (positive or negative) and in intensity, whereby the intensity of the valuation is relative: considering an object of attitude on its own involves implicit comparison to a set of objects perceived by the evaluator to be of the same kind (Kahneman et al. 1999). As Kahneman and colleagues observe (Kahneman et al. 1999), “objects of attitudes [as the term is used psychology] include anything

that people can like or dislike, wish to protect or to harm, to acquire or to reject.” Henceforth, we shall say that an expressive communicates *evaluations* to cover the evaluations that arise from attitudes, emotions, moods, or feelings. We shall, however, look into these last concepts in more detail later on, in the Chap. 4.

- *Declarative*. A declarative speech act brings about the conditions specified by its content, provided that the role of the speaker allows the realization of the conditions. A president announcing war and a judge issuing verdict both use declaratives, thereby bringing about the conditions conveyed by the speech act. The content is believed to be true by the speaker.
- *Representative declarative*. A representative declarative speech act is used by the speaker to acknowledge that the conditions – those described by its content – hold. As for the assertive and the declarative, the content of a representative declarative speech act is believed true by the speaker.

Assertives, declaratives, and representative declaratives communicate beliefs of the speaker, directives convey desires, commissives intentions, and expressives evaluations. It is essential in an analysis of advice to distinguish between the conditions that actually hold and those that are desired, but do not hold when advice is given. Any ontology of advice should distinguish advice that suggests what should be believed, from that which says what should be desired, intended, or what evaluations should be had. An athlete may be advised to believe that a new world record result is breakable, whereby he is advised on what to believe. He may then be advised that he should desire to break that world record, which is different from believing it can be broken. Further, if he is advised on what to specifically do to break that world record, he is advised on what intentions to adopt. Finally, he may receive council that he should prefer breaking the world record than merely winning, in which case he is advised how to evaluate the potential effects of his actions. It should be self-evident that these different kinds of advice are not identical, in the sense that they may have different effects on how the recipient of advice will decide after he receives advice. Advised desires may orient the individual in his search for alternative courses of action, and thus influence the range of intentions he may consider before acting. Advised intentions instead already say how to act, thereby aiming to directly influence the range of alternative courses of action the individual may choose, regardless of his desires, evaluations, or beliefs. Advised beliefs, as in, say, religious morals, disregard desires, intentions, or evaluations, aiming to tell what is possible or socially acceptable. If the athlete believes he cannot break a world record, then this could effectively influence the desires he may have (e.g., being among the first five or ten, instead of aiming for the first place), the evaluations he may form (e.g., it is better to train for a first-five place, than risk burnout by training for the world record), and his intentions regarding, say, training, nutrition, and so on. Consequently, the fundamental difference that any ontology of advice should make is that between beliefs and desires that are advised. In order to do so, it is not necessary to finely separate assertives from declaratives, for they both convey beliefs. Instead, it is essential to distinguish assertives, declaratives, and representative declaratives from commissives: the former convey beliefs, while the latter convey desires. We will argue below that *evaluations*, which are conveyed

by expressive speech acts, require a treatment distinct from that of beliefs, desires, and intentions. Thus, the scope of an ontology of advice will have to encompass distinctions between advised beliefs, desires, intentions, and evaluations.

To assume that communication underlies all advice, that advice cannot exist without, or is transferred through communication, effectively gives the limits of the scope of our advice ontology. It starts from a simple observation, *prima facie* evidence: something cannot be called advice if it has not been communicated. To dispense advice, an individual must communicate the content of advice. In doing so, he will communicate this content in some particular way, using one speech act or another. To distinguish then dictum from modus allows a classification of advice to be made, according to the way it has been communicated, or rather, the modus that its recipient associates with the content of advice.

### 3.4.3 *Essential Properties of and Identity Criteria for Advice*

The previous section identified the first pair of essential properties of advice. As anything called advice must be revealed through communication, it has both a content and a mode in which it has been communicated.

People speak of dispensing or giving, and of receiving advice. At least two roles are consequently involved in any communication that involves giving and receiving advice. Although it may seem strange to advise oneself, even if this is admitted, the same individual will not play the same role when giving and when receiving the advice he is dispensing to himself. We can picture Selkirk, in a moment of forgetfulness or folly on his desert island, saying to himself that he should not do so and so. Despite the sun or solitude having hit him hard enough to warrant his speaking to himself, he still cannot avoid changing roles from the advice-giver to advice-recipient when he dispenses advice. This necessity for two roles to be involved in the communication of advice matters, because it says that advice requires both the sender, and the recognition by a recipient. Advice is thereby something social: some content communicated in some mode will need to be recognized as advice at least by its recipient. By being something social, whatever is wholly confined to the mind of a single individual cannot be advice. An object solely within the private bounds of a single mind would be some mental object, inscrutable to other individuals, except of course in the rather unlikely event of mind reading, for now in science-fiction only.

That any communication that is potentially also advice has the property of being something social, that is, of being recognized by more than the individual that originates that communication, begs the question of intentionality: must the individual recognized, by the recipient of advice, as the giver of advice have the intention to give advice when he communicates for that communication to be called advice? While the affirmative answer may have been expected, it turns out a failure. We have repeated on several occasions above that mind-reading is not an option. Consequently, when the recipient is convinced that the giver did indeed have the intention to advise, this conviction does not mean that he is right: namely, that the giver



dispensed advice because that giver had an intention to dispense advice. The recipient can ascertain his conviction only if he could read the mind of the giver. As this cannot happen, the recipient's conviction arises not of an ascertainable existence of the giver's intention, but from the recipient's formation of reference relations. The recipient establishes himself, independently of any potential intention of the giver, some reference relations between signs – the *prima facie* evidence – he observes and is communicated to in the advice-giving context, and his very own concept of intention. The recipient may thus claim that the giver intended to advise, yet such a claim can only be seen as an imprecise and misleading explanation, for nothing can be known of the intention. Of course, the recipient may ask the giver if the latter advised intentionally. Even if the giver says he intended to advise, this does not mean that *any* advice must involve an intention to dispense advice from the giver. The advice-giver's intention to advise is therefore not an essential property of advice: as long as the recipient understands the content and mode of a communication as being advice, no need for him to assume further that the advice was dispensed intentionally. Note the consequence of not having intentionality of advising as an essential property: it is enough that someone hearing or otherwise experiencing a communication considers it as advice, for that communication to satisfy one identity criterion of being called advice.

Advice is a nonphysical object. As opposed to physical objects, such as chairs, people, books, and cars, advice is like a piece of music, a convention, a principle, or a law. It can certainly be produced by and recorded onto physical objects, but even so it remains distinct from them; e.g., we may make a copy of the nonphysical object on another physical object, and even if the copy fails to be perfectly similar to the original, it can still be recognized as resembling the original nonphysical object, as its copy. Advice is nonphysical in the same sense principles are nonphysical, such as “form ever follows function” that Louis Sullivan, a nineteenth and twentieth century architect and Frank Lloyd Wright's mentor, coined in his essay *The Tall Office Building Artistically Considered*. He tellingly separates the physical from the nonphysical in arguing for his ground principle of modernist architecture. As the price of steel was falling throughout the nineteenth century, it became possible to change the way tall buildings were constructed. Height was limited by the thickness of the walls, since these had to support the weight of the floors. With a skeleton of steel girders, to which all other elements of a building were to be suspended, the tall office buildings, and later skyscrapers had become feasible. In arguing that form ever follows function, Sullivan was arguing that social conditions govern the function, and that function governs form, in place of, say, artistic trends of the moment:

[O]ffices are necessary for the transaction of business; the invention and perfection of the high-speed elevators make vertical travel, that was once tedious and painful, now easy and comfortable, development of steel manufacture has shown the way to safe, rigid, economical constructions rising to a great height; continued growth of population in the great cities, consequent congestion of centers and rise in value of ground, stimulate an increase in number of stories; these successfully piled one upon another, react on ground values; and so on, by action and reaction, interaction and inter-reaction. Thus has come about the form of lofty construction called the ‘modern office building.’ It has come in answer to a call, for in it a new grouping of social conditions has found a habitation and a name. (Sullivan 1896)



It is the physical form, the building itself that is the physical object, shaped in Sullivan's view by the nonphysical, the social conditions such as the necessity to perform business in some desired manner. With form that follows function, the nonphysical object, beyond recording, shapes the physical one. Returning to our preoccupation with advice, an advice – say, that form should follow function – is certainly a nonphysical object.

Is it enough for any social nonphysical object that is communicated, has a content and mode, to be advice? Anything communicated is a social nonphysical object and has a content and mode. These are consequently not sufficient identity criteria, merely the necessary ones. The distinguishing property, one that when added will be enough to single out advice among other social nonphysical objects that arise out of communication, is relevance to the recipient's decision problem. As we said earlier, the recipient must be able to establish the reference relation between the objects that are accessible to him and the signs used in that communication, for that communication to be called relevant to the recipient's decision problem.

This brings us to the following intensional definition of the concept *advice*:

**Definition 3.1. Advice:** Any instance  $x$  of the concept *advice* must satisfy the following identity criteria:

1.  $x$  is some potentially complex speech act that has been performed an individual; we will call that individual the **advisor**.
2.  $x$  has been experienced by an individual; we will call this one the **recipient**.
3. The recipient can distinguish the dicta from modi in  $x$ , and from the modi establish which of the dicta he could adopt as beliefs, desires, intentions, or evaluations.
4. The recipient can form reference relations between at least some of the dicta in  $x$  and objects in his context of reference.

Several remarks are in order on the above definition of advice:

- *Wordnet*, the lexical database of English, says that advice is any recommendation on a course of action. As defined above, advice seems to have a wider scope. If the definition above should be restricted to recommendations on actions only, then we could only accept those speech acts that the recipient refers to his own potential actions. That is, communications that the recipient may integrate as new or changed own desires, beliefs, or evaluations would not be taken as advice. Now, we only say that the scope *seems* wider in our definition of advice, because the *Wordnet* definition can itself be read in a very wide manner. For example, the act of adopting something as a belief can be called a course of action, so that a speech act that the recipient understands as suggesting a belief to adopt can fall under the *Wordnet* definition. If we went the other way around, aiming to obtain an informal definition of advice from the definition suggested above, then advice would be anything that an individual is told (or communicated otherwise), and that he sees as potentially believable, desirable, doable, or as evaluations (e.g., preferences) he could apply himself. However, one reads the *Wordnet* definition and the definition above, they certainly do not contradict each other.

- As we argued earlier, advice cannot happen without communication. The above definition reflects this observation by involving two individuals, or two actors, one of which is giving advice, while the other one is receiving it.
- As speech acts are nonphysical and social objects, the definition ensures that any advice is a social nonphysical object.
- The third identity criterion speaks of dicta and not dictum, and modi in place of modus. Plurals are used, as it is impossible to ascertain that any advice will be communicated by primitive speech acts. The advisor may use complex combinations of speech acts. The definition says nothing about how these are to be decomposed by the recipient, and how he is to distinguish the dicta from modi. These processes are not very clear in general, and this definition cannot clarify them.
- The fourth identity criterion says that dicta are signs. The recipient of these signs must form reference relations between these signs and the objects in his context of reference. This identity criterion is due to the earlier discussion of relevance: if the recipient cannot refer the signs to objects in his context of reference, the dicta of advice is irrelevant.
- The definition allows various lines of specialization to be followed. As there are dicta and modi in advice, specialization can be pursued along the properties of the content and/or of the mood of advice. One way to specialize can be to separate advice that refers to some specific classes of objects, from other. Doing so would amount to specialize advice according to the properties of its dicta. As some speech acts issue commands, while others convey beliefs, modi can straightforwardly be used to classify advice. With the third identity criterion saying that the recipient decides what he adopts as beliefs, desires, or otherwise, advice can be classified depending on how the recipient classifies it within his beliefs, desires, intentions, and evaluations. Some of these lines of specialization will be explored below.

#### 3.4.4 *What Advice Is Not, but May Refer To*

Identity criteria for advice tell us unambiguously what advice is. That, at least, is what would be hoped from any definition. Beyond hope alone, it is relevant to say what advice is not. Doing so makes further ontological commitments explicit, those that are not directly apparent behind the very definition of advice, thereby placing the definition offered earlier in a wider, foundational ontology.

Admitting that advice is a nonphysical and social object begs the question of what other objects there are besides nonphysical social ones. Social objects should be distinguished from those that require no social recognition to exist, and which are also nonphysical. These can be called *Mental objects*, being confined to the mind of the individual and existing within these bounds. The separation between social and mental nonphysical objects is a complete partition of the class of nonphysical objects. That this is the case can be accepted only if it is taken seriously that the

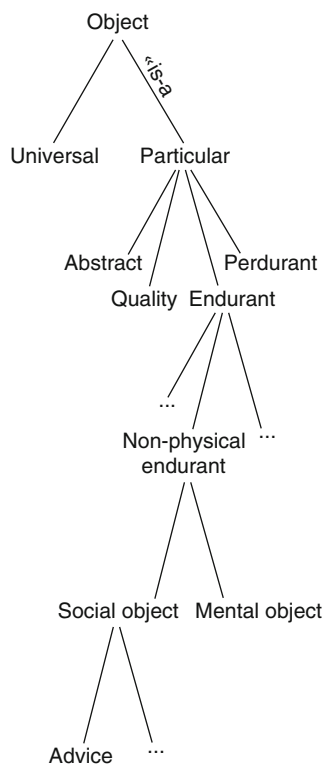
existence of any nonphysical objects crucially depends on people, or rather, of any being that can exhibit beliefs, desires, intentions, and so on. If there are conventions between animals, then these are also social nonphysical objects. However, as soon as there are no intentional individuals, there are no nonphysical objects.

To further clarify the classification of advice in terms of foundational concepts, an answer is needed to what nonphysical and physical objects are specializations of. Are nonphysical objects specializations of *events* we spoke of earlier? Advice is produced through a process, which involves the performance of speech acts. The speech act, its content, and mode must be recognized in a particular way, according to our definition of advice; the recipient must form reference relations between signs he has been communicated, and the objects in his context of reference. To the extent that advice integrates the recipient's beliefs, desires, intentions, and evaluations, it exists in time, ceasing to accumulate parts over time. That advice can be received and given, bought and sold, yet retain identity leads to conclude that it cannot be a perduring object, but an enduring object. The latter are wholly present after they have been created, pending destruction or damage so extensive that the damaged object no longer satisfies identity criteria.

Advice is thus a kind of *endurant*, a nonphysical social *endurant*. Being an *endurant*, any instance of advice is a particular, but not a *perdurant*. We thus reach the class of particulars, which we mentioned in speaking about the specialization of the object concept, and in Fig. 3.3. But is any particular either an *endurant* or a *perdurant*? If not, then we can say more about what cannot be advice. Namely, it can be said that by being an *endurant*, advice is neither a *perdurant*, nor a quality, nor an abstract. Any specific advice, what can be loosely called a piece of advice, is not a quality of either a physical or a nonphysical *endurant*, or of a *perdurant*. The term *Quality* refers to anything we can perceive or measure on any object, such as color, shape, size, weight, and which are inherent to objects they are perceived or measured on: a quality of an object cannot exist without the object. Finally, while some specific advice may refer to sets, or scientifically established facts (e.g., that Earth has spherical shape), it cannot be equated with these (instead, advice may refer to them). Facts, sets, and regions are usually considered as abstract particulars. Abstract particulars differ from *endurants*, *perdurants*, and qualities in which they are not qualities themselves and have no inherent temporal or spatial qualities. If we consider the fact that Earth has a spherical shape, this is not something that is rationally justified because someone did an observation to establish it, but is, for all practical purposes, something that has been true regardless of such observations taking place, that is, before, during, or after any observation aiming to question the truth of that suggestion.

By placing advice among *endurants*, *perdurants*, qualities, and abstracts in Fig. 3.6, a foundational ontology must have been hypothesized in the discussion above. That foundational ontology is one of particulars, where the class of particulars is the one being specialized onto *endurants*, *perdurants*, qualities, and abstracts. The *endurant/perdurant* split is an usual one. It seems to fit the distinctions that people make when speaking of events, processes, and, say, physical objects, laws, principles, and so on. When spatiotemporal qualities matter in no way, abstract

**Fig. 3.6** Position of the concept of *advice* within a multiplicative and descriptive foundational ontology



objects enter the picture, although without falling within either endurants or perdurants, being thus a separate class of particulars. As we clearly can speak of colors, sizes, weights, and other properties of specific particulars, the foundational ontology requires a class that would encompass these. Hence the class of qualities. Beyond these distinctions, or rather, before them stand the very first ontological commitments. One among them is the decision to have either an engineered or a metaphysically ambitious ontology; given our prior discussions, we clearly chose to go for an engineered foundational ontology. This can also be stated otherwise: the foundational ontology here is a descriptive, not a revisionary one. A descriptive ontology hopes what can loosely be called the ontological categories underlying natural language and human common sense. In practice, this laudable aim translates into choosing to specialize particulars onto categories that seemingly go not too far away from the basic distinctions people seem to be making with a natural language. It seems to be *prima facie* evidence that people do see events as different from, say, physical objects. With a descriptive ontology, metaphysical ambitions of revisionary ontological commitments disappear for the aim is not to somehow capture exactly what may truly exist.

Another choice for a foundational ontology is if it would be reductionist or multiplicative. One way to distinguish between these two, and see which may be

preferred, is to consider how someone who commits to either of these kinds of ontologies would speak. In a multiplicative ontology, he would say that *the paper flight ticket is constituted of an amount of paper*, whereas in a reductionist ontology, we would say *the paper ticket is an amount of paper*. A multiplicative ontology allows different entities to be co-localized in the same spacetime. In the former example, a paper plane ticket coexists with the amount of paper it is made of, occupying the same spacetime. We can say the ticket is a social nonphysical endurant, while the paper it is made of is a physical (and not a social) endurant. It seems less contentious to relate the paper of the ticket and the ticket by saying that the former constitutes the latter, instead of claiming that the latter *is* the former. Commitment is thus on a multiplicative account, not a reductionist one. The choices of how to partition particulars, and to adopt a descriptive and multiplicative stance are not new. Nicola Guarino and his colleagues' proposal for an engineered foundational ontology partitions the class of particulars onto endurants, perdurants, qualities, and abstracts, and the discussion above has in this sense been inspired by theirs (Masolo et al. 2003).

### 3.5 Relativist's Conceptual Analysis

Two questions were asked at the outset of this chapter: (1) *what is the advice referring to within the context of the specific decision situation?*; and (2) *what is the of purpose of that advice in that context?* Conceptual analysis of advice offers guidelines to follow in responding to the first question.

To know what advice is about, and hence to be able to act on advice – at least decide whether to accept it – the recipient of advice needs to have had, or have access to a series of definitions, of the signs used in advice. When an individual is advised that he ought to carry an umbrella for it will rain tomorrow, he is only able to act on this advice if he has some idea – ideally very close to that of the advisor – of what an umbrella is, of the consequences of it raining, and so on. This is all very simple in many cases, but once we step out of the apparent everyday matters and step into the realm of, say, politics and economics – in general, into elaborate systems of coordination – matters become considerably less straightforward. It becomes a challenge for the specialist-turned-generalist to know what some advice he is given is precisely about, or in other words what it refers to and what consequences its acceptance or rejection may have.

In order to understand why it is so difficult to design relevant definitions and thereby relevant advice, we looked into the notions of *sign*, *concept*, and *object*, and the *reference* relation. Putting them together in the sign/concept/object triangle was a way to highlight why good definitions are hard to come by and why it is difficult to ensure that advice is understood as the advisor intends. Every definition relates a sign to what it is supposed to stand for, that is, a definiendum to a definiens. Both the definiendum and the definiens are always only signs, which are then interpreted by the user of the definition. Same goes for advice: any advice is a collection of

signs, which are then to be interpreted by the recipient. Meaning, however, goes through reference, and it is, so to speak created via the mechanism of reference. The reference relation in turn forms within a context, and with the varying of the context, the same sign need not always refer to the same concepts and/or objects. Signs and the elements of context determine reference.

The lesson an engineer of advice should learn from the knowledge of what signs, concepts, and objects are, and how they relate is that signs used in advice should be chosen on the basis of (1) what concepts/objects advice should refer to, and (2) assumptions about what will form the context, in which the advice will be delivered. The latter assumptions include also hypotheses about what the recipient of advice may or may not already believe, desire, intend, their emotions, feelings, moods, and so on. The lesson that a recipient of advice should learn is that when he analyses the advice he receives, his focus should turn to what signs in advice and what elements of the context, in which he received advice influenced how he interpreted the advice that he had received.

Only if the analyst and engineer of advice understand how, roughly speaking meaning gets created, can the former attempt to design advice before giving it, and the latter can say that he can analyze it before deciding whether to accept it.

As reference relations cannot be established once and for all, but are context-dependent, conceptual relativism becomes a plausible stance: people construct over time their personal conceptual schemas, they have their own language and concepts to classify particulars and organize them into some picture of the events they are taking part in, that is, their own understanding of it all. Facing conceptual relativism, we are confronted to the difficult problem of how to share information about various conceptual schemas, precisely to reach agreement, or rather, reduce disagreement to some satisfactory extent. It is effectively the problem of the engineering of advice, that is, how to combine various available signs to influence the formation of reference, i.e., control to some extent the understanding that recipients will have of advice. This led us to the engineering of ontologies, as a way to render parts of conceptual schemas explicit, so that they can be shared and discussed.

What was argued for in this chapter was subsequently applied to arrive at a definition of the term *advice*. To see why conceptual analysis does matter, beyond perhaps the obvious, consider why it is relevant in the first place to have some definition of *advice*. Clearly, the very basic purpose of any definition is the classification of phenomena, along with the agreement on that classification: an individual who knows the definition of advice can recognize something as being advice (i.e., that he can establish the reference relation between the term advice and some particular communication). It is moreover evident that we cannot hope to develop an analysis of something if we fail to agree what properties it seems to have: an individual who wishes to develop techniques for the analysis of advice needs to understand what properties the object of analysis has. Finally, an individual who wishes to design advice needs to understand what essential properties must be satisfied by the product of his engineering effort. Not only is then an explicit definition needed to agree as much as feasible on what we are talking about here, but also because it influences ideas of what any analysis of advice is supposed to be analyzing, and what

any attempt to design advice aims to produce. Exactly the same remarks apply to any sign used in advice.

Despite the danger to vulgarize, it is still worth repeating what broad steps were taken in the conceptual analysis of the term *advice*. The analysis starts with the simple question of what makes a definition, to which a reply of definiens and definiendum is not good enough. As soon as it is recognized that both the definiens and definiendum are signs, it is relevant to place the term *advice* as the sign in the sign/concept/object triangle, which leads to two questions: (1) what concept does the sign *advice* refer to, and (2) what objects does that same sign refer to. Candidate objects are any particular advice, while the concept should carry properties that *any* particular advice exhibits.

These questions initiate *the second step* in the analysis, namely, an imprecise account of the properties that seem to be shared by various particular advice. This is basically a description of a conceptual schema of advice. From *prima facie* evidence about advice, that is, from a conceptual schema about advice, it was argued that there can be no advice without communication. Some information can only be called advice if it is communicated. An essential property then of any advice is that it is communicated information. This in turn begs another question, namely, is *any* communicated information advice? The answer we argued for is the negative, because the communicated information must somehow be relevant for the decision-maker who is being advised. Interest ensues in what would be a criterion for such relevance: what condition should some communicated information satisfy to be called advice? The answer to this question would be another essential property of advice. The answer we argued for is that it must be possible for the recipient of communication to establish reference between the signs in that communicated information and the elements of his context. If the communication was such, our argument went, it effectively satisfies two essential conditions: (1) the information is communicated, and (2) it is relevant to the context of the recipient. From there on, we had two essential properties of advice.

*The third step* that ensued aimed to clarify – explain may be too strong a word – what these two essential properties refer to. To get to some clearer understanding of communication, we looked at John Searle's speech act theory. The picture this gave is that advice is produced by the performance of speech acts, actions that the advisor does to influence the beliefs, desires, intentions, and so on. Advice thus involves necessarily at least two people, as the definition highlighted, the advisor and the recipient of advice. Another equally important idea gained through speech acts is the useful, though not sharp separation between the content of a speech act and its psychological mode, i.e., the distinction of what is said from how it is said. This separation, as is visible from the suggested definition of *advice*, allows the definition to incorporate the other essential property of advice, namely that the recipient can form reference relations between at least some of the dicta in *x* and objects in his context of reference.

The consequences of the ontological choices made throughout the first three steps, and the ensuing definition are significant. Any attempt to engineer and analyze advice must be concerned with what is communicated in advice, how it was

communicated, and in what context that exchange took place. The *final step* of the conceptual analysis was to make it clear what advice *is not*. This was achieved by clarifying the position of the concept of advice within a foundational ontology, which itself makes ontological choices compatible to those we made in defining the term advice.

### ***Homo Follis, Part Two***

*Ontological choices* are simply a technical name, a specialized sign that we use to refer to actual choices that an individual makes when he adopts some signs, concepts, objects, and reference relations to organize his thinking about whatever he cares for. The objects of one's interest may well be particulars or universals, phenomena or abstractions that have some role within his conception of self, others, and the various surroundings. Making sense, explaining both one's own and others' behavior, and the things and events in the surroundings involves the use of signs and reference relations between the many and various particulars and universals. To the extent that such explanations influence his future behavior, ontological choices do affect future behavior. When – according to the anthropologist Claude Lévi-Strauss – a priest in some seventeenth century Peruvian tribe sees the weather getting much too cold, he summons those who have been born feet first, or had a harelip, accuses them of causing such weather and orders them to repent (Lévi-Strauss 1995). The priest's act is no accident: to say that it is grounded in myth is another way to say that the priest's conception of the event in question and of its causes is somehow misguided, as science says that there is no link between being born feet first and causing weather to get cold. The priest operates on assumptions, ontological choices made precisely to introduce some order in what may appear as chaos. The myth itself incorporates ontological commitments, the explanations for actions arise out of the ontological commitments, and action itself is performed as the indirect result of having made specific ontological commitments. The modern president urging for hope, or the other urging for invasion operate in the very same sense on grounds of ontological commitments; differences lie in how well accepted or convenient these ontological commitments are to those bearing the effects of acts that these commitments produce and rationalize.

The influence of ontological commitments on acts is indirect. The objects of experience and intuition are made sense of, are explained to oneself through complex combinations of signs, concepts, objects, and reference relations. Ontological choices enable such explanations, they stand as a basis on which explanations are erected. They justify the complexes made of many signs, concepts, objects, and references.

Insofar as acts that are not accidental are the result of thinking, these acts are a reflection and result of the intellectual construction standing in their shadow, and this regardless of how elaborate or trivial that construction may be. When an observer of the purposefully acting individual is to make sense of the observed act, he will seek



to rationalize the act and its effects. He will thus be obliged to seek the ontology that the acting individual engineers, and within which the act obtains its meaning. The role of the conceptual analysis of advice is precisely to ask questions about that engineered ontology, the collection of related concepts that the acting individual manipulates to rationalize own acts. To perform the conceptual analysis of advice with conceptual relativism in mind is to reject the metaphysical questions of what definite and universal ontology may stand behind the engineered ones; it is to reject the idea that acts obtain some perspective-independent and actor-independent meaning within that universal ontology. Rather, by seeking to elucidate engineered ontologies within which acts obtain a justification, the conceptual analysis of advice very much accepts that stories behind acts may rely on specific languages, that the act, the story of the act, and the language of the story are all interdependent. The very purpose of the analysis is to elucidate the specifics of a language, and thus of the underlying ontological commitments. The specialist-turned-generalist, who applies the conceptual analysis of advice to the recommendations he is offered, effectively asks what the story of these recommendations is, what is the language behind it. To answer these, he must ask what the engineered ontology is, and thus what ontological commitments stand in the shadow of the available recommendations. The analysis moves backwards from the available advice, to its rationalization (the story, for lack of a better word), on to the engineered ontology (in terms of which the story is built), and finally the ontological commitments out of which arise the categories of particulars and universals recognized in the language.

This chapter thus needs two readings, one following the page order, and a reverse one: from the recommendation on how to define the term *advice*, to the engineered ontology of communication and reference behind it, to what engineered ontologies themselves are, and finally to the sign/concept/object triangle. Each step of the way, from the use of a sign in advice to the ontological commitments given to justify that use, the acceptance of advice is the question of the extent to which the analyst's and the advisor's stories, engineered ontologies, and ontological commitments depart from and rejoin one another. To perform the conceptual analysis of advice is to ask at which points available advice meets one's own conceptions, and what the consequences may be in cases of departure.

Not to be concerned with the conceptual analysis of advice is either to fail to grasp the nuances in and behind the advice one designs or analyses, or to decide whether to accept advice by flipping a coin.

## Chapter 4

# Interpretation of Advice

A photograph shows a woman's face out of focus; she holds her index finger in front of it, pointing up. Her finger is covered from its tip onwards up to about one-third of its length in indelible black ink (Marai 2009). Her dignified expression is closest perhaps to indifference. The face is carefully veiled. Her brown eyes look straight into the camera. She is distinctly Asian. According to Agence France-Presse, the photograph was taken after this woman cast her vote on the elections in Afghanistan on 20 August 2009. The ink is there to ensure that she cannot cast another vote at another polling station. What does this photograph refer to?

There is the immediately apparent: it is a veiled woman of Asian origin; she is holding her index finger in front of her face; the finger is covered in black ink. There is nothing to incite doubt or wondering about the immediately visible. Although the objects captured on film are clear and shown in a rather unsurprising arrangement, the reader of a newspaper featuring this photograph will have to rely on much more than what the vivid colors and the elegant shapes convey alone. Much of what is salient and available, much of the original context is removed. Nothing remains of the ambient noises and sounds, of the interaction between the photographer and the subject prior to the shot, the behavior of the people surrounding the photographer and the subject at the time of the shot. If nothing is said of where and when the photograph is taken, few could establish from it alone that it has indeed been taken in Afghanistan, and on that particular day. There is little in the picture itself in terms of contextual cues that would without doubt exclude it being taken in another country, where polling stations also use black ink. The extent to which the twentieth century has rendered traveling inexpensive makes it impossible to deduce from the appearance of her face that the picture must have been taken in one particular country, and not another.

Although a picture alone may say thousand words, apparently a thousand is not good enough. When a *The Boston Globe*, a newspaper, featured it in its digital edition (Taylor 2009), the caption read: "An Afghan woman displays her finger marked with indelible ink after casting her vote at a polling station in Kabul on August 20, 2009. Afghans voted to elect a president for just the second time in their war-torn history as a massive security clampdown swung into action to prevent threatened Taliban attacks derailing the ballot." It is with the caption that we can go beyond the immediately apparent reference in the photograph. And it is this specific caption

that tells us that we should indeed go very far. The second sentence in the caption does most of that work. Without the caption, all signs from which to form reference are only those within the picture. Add the first sentence, and doubt about places and times disappears. It is with the second sentence that the picture can obtain a symbolic status. The second sentence opens up the possibilities for reference far beyond what can effectively be seen in the picture itself, and known after the time and location are given.

What reference is formed by looking at that photograph certainly does depend to a significant extent on the contextual cues, that is, the signs within and given together with the photograph. To think, however, that it depends *only* on the contextual cues *other* than the viewer himself would be a fallacy. The journalist can work backwards from the message to communicate to the selection of signs, including the picture and the surrounding text and graphics. All of these can be carefully put together to restrict reference. But the thousand words are only so good, for the beliefs, desires, intentions, and evaluations of the viewer are a significant part of the contextual cues that affect the formation of reference. These have been formed through the interminable drill of many pictures, along with all other outputs of various coordination mechanisms, including media, political, and economic systems. Not to deny the autonomy of the individual, we should certainly give credit when distance is taken, and thinking engaged to counter passive reception. Both of these are obvious, but are nevertheless important to repeat here, for it is they that ultimately shape reference. To return to the photograph, how the caption reads may well be credited to the journalist who intended to hail the arrival of something called democracy to a country called Afghanistan. The photograph should in this approach serve as evidence to the reader, of the progress made despite the horrors of the local conflict from the outset of the twenty-first century. It would tell that choice has been given to the people even there, and that a woman in a highly conservative country has made her choice, marking thus her autonomy. Going further, though the finger she points is the index, it is not too far off to see this as some cultivated variant of giving the finger to whatever force may be against letting her decide, partake in the democratic process. Such a reading is not far off precisely because the caption reminds us of the menace against voting in Afghanistan, given this country's traditional ways to transfer political authority.

There is no need to change any of the signs surrounding that photograph, to make very different references. It is enough to have a slightly more elaborate knowledge of the matters in Afghanistan, and approach reference with at least some skepticism. In other words, instead of changing the elements of context that are given independently of the viewer, change some of the viewer's beliefs, desires, intentions, or evaluations.

An astute viewer of this photograph and of the material accompanying it may recall that the British empire had no other way to keep its superiority in the nineteenth century but to maintain its hold on British India and pre-empt any further incursion of Russia into Central Asia. According to Edward Ingram, a historian:

The first industrialized state and the first free society, in its own eyes naturally, was to take advantage of its superior technology, its steam power, its iron and its cotton goods to take

over and develop the economy of Central Asia. And after British goods would follow British values, in particular, respect for private property. Given security for the just rewards of labor, nomads would settle and oasis cities surrounded by tribes of herdsmen would be turned into territorial states with agreed frontiers on the European model. (Ingram 1980, pp. 164–165)

The First Afghan War thus involved the establishment of a puppet government with a limited military support from the British. The new ruler would in return open Afghanistan to British trade, and while the new government was put in place in 1839, it could not remain despite Britain's attempts until a retreat in 1842. After this, two other Afghan Wars were waged, the last of which took place in 1919 and ended Britain's influence in the Afghan foreign affairs. The country spent the rest of the twentieth century trying to remain neutral, and did so in the Second World War, while oscillating between revivals of religious tradition (or extremism, depending on who is asked) and attempts at modernization, the latter through compulsory (mixed) education and the abolition of the veil for women, and their later involvement in politics. The last failed attempt at modernization in the twentieth century was by a communist government, intent on improving education, the position of women, and the freedom of religion. This, along with destabilization from the Soviet Union and the USA set off a civil war in 1979. Between 1979 and 1989, when the Soviets withdrew from Afghanistan, up to two million people had been killed in the conflict. The civil war continues at the end of the first decade of the twenty-first century.

Even a limited knowledge of the Afghan recent history awakens skepticism, and hardly allows a simplistic reading of the photograph and signs packaged with it. When shown to the uninterested nonskeptic, it could refer to the victorious march of democratic progress, toward some renewed Afghanistan that is but a stone throw away from some nearby polyarchy. It would be ironic, but not unexpected if the British reported something similar to their readership in the aftermath of the 1839 transfer of power. To the interested skeptic, the least benign reference this photograph could produce is that it simply fills up the media space. While there clearly is quite some junk within that media space, proper readings are unlikely to be so favorable. Is this picture not given and precisely set as is it to say that the military intervention was a necessary evil, but that all that now stands in the shadow of progress? Does it not glorify individual choice in a country which apparently allows very little of it? Is it so implausible to think that the elegant lady actually is giving the finger, not to the traditionalist oppressors within her country, but to the viewer-democrat from the outside?

The multiplicity or openness of reference is not a problem. It is rather a solution, a safeguard against the usually present and variously pronounced idiocy of those producing the signs. If there is something that a dictator should wish for, it is that signs carry over the exact reference that their producer formed, without distortion or openness. This would be to liken speech acts to surgical instruments for a brain operation: while the latter manipulate pieces of brain tissue, the former would manipulate the beliefs, desires, intentions, and evaluations of the recipient. In doing so, they would be considerably more dangerous. The sender could tailor the signs in the content of speech acts and choose the mode carefully to obtain some desired effect on the attitudes of the recipient. This twisted prospect nevertheless fails miserably. While

the surgeon has extensive control over the effects of his instruments on the tissue, the individual producing the speech acts – giving advice – is confronted to the countering attitudes of the recipients and their varying degree of autonomy in choosing what to accept and what to reject. Hence perhaps the use of incentives, or in the case of detainees, of physical and mental torture, speech acts never manage to go all the way.

Any individual who receives advice forms reference from the content and mode of communication, but also own beliefs, desires, intentions, and evaluations. Just as when reading the said photograph and its surroundings, the individual is to form reference from the signs used to communicate advice. The variety of reference that the picture allows are not merely a metaphor for the variety of references advice allows. There is effectively no metaphor: the problem is the same. The picture we spoke of provides itself advice: it is shown in order to advise on beliefs, desires, intentions, and evaluations an individual may hold regarding the then-conditions in Afghanistan.

Given the apparent difficulty to pin down the reference that signs will evoke, we will discuss further the openness of reference in this chapter. Since the information and attitudes held by those receiving advice plays a significant role in the formation of reference, we will look into kinds of advice in the next chapter, depending on what it targets – the beliefs, desires, intentions, or evaluations of the recipient. In doing so, we will specialize the concept of advice defined in the previous chapter.

## 4.1 Open Reference

Already formed reference is open in the sense that it is very rarely independent of new information that may become available. New information act as new contextual cues, potentially changing the previously established reference relations.

A convenient illustration to how significantly reference can change with the arrival of new information are the various photographs that occupy a cult status in modern culture, including Robert Capa's "The Falling Soldier," Mathew Brady's American Civil War photographs, and Lewis Hine's records of the industrial development in the USA of the early 1920s, among others. Philip Geffer, a writer about photography, notes in an essay for *The New York Times* that the optical precision of the photograph seems to promise a truthful reflection of the world, an authentic recording of events. Yet he rightfully argues the obvious: "just because a photograph reflects the world with perceptual accuracy doesn't mean it is proof of what spontaneously transpires" (Geffer July 23, 2009). An image's status as evidence comes only out of it being called evidence by some authority that claims the veracity of that image.

The photograph known as "The falling soldier" (also as "Loyalist Militiaman at the Moment of Death, Cerro Muriano, September 5, 1936") shows a man collapsing to his back, holding a rifle in his right hand. He looks as if he has been shot in the head while posing, and the photographer managed to capture the very moment

of death. The picture is among the best known of the Spanish Civil War. As much of war photography, it evokes the tragedy and the senselessness of violence, here perhaps in the most striking manner, as the very moment of death is evoked. This of course means that there is more in terms of reference here than merely what the photograph captures:

It is an emblematic image: it suggests that individuals have ideals and are prepared to die for them. It proposed that war remained the arena of individual honor and bravery, and that even paying the ultimate sacrifice furthered the cause. [...] [it] contributes to a purposeful sense of war as sacrifice. This notion is required to sustain high morale, especially among civilians looking at illustrated papers: it depends on officially approved photographs in which (for instance) soldiers prepare, go forward, capture the enemy and sometimes suffer quick, clean deaths. (Taylor 1999, p. 161)

We can of course read in it something rather different upon learning that it has been staged, that the soldier most likely did not die at that moment, as has been repeatedly argued. It then no longer has its prior gravity. Instead of being a symbol against violence, it becomes for some the symbol of manipulation, perhaps even ridicule of a grave period. For others, it remains a depiction of, say, sacrifice in war, even if a slightly adjusted one; a work of art, not of investigative reporting. In both cases the initial reference changes: the original signs are complemented with those that indicate staging, leading to qualify the photograph differently, that is, seeing its qualities as instantiating different properties. The photograph becomes the sign referring to an instance of a different concept. If the photograph was indeed staged, it can no longer be a sign of the presumed event, namely the death of the depicted soldier. The sign has lost the object it has been the reference of and becomes the sign of another event, namely that of the staged death, which still did take place on the same time and at the same place as the prior object of reference.

The discussion of the Falling Soldier illustrates two important observations. Firstly the same photograph is a sign referring to potentially many different objects and concepts. At the very least, it refers to an event that took place on 5 September 1936, and to the concept of sacrifice in war, which it can be said to symbolize. Second, knowing whether it is a staged photograph that leads to the formation of different reference relations, in which the photograph acts as the sign in the sign/concept/object triangle. The first point is that multiple reference can be present, and is actually likely to be given that reference depends on the individual and contextual cues in addition to depending on the sign itself. The second point is that the reference relation need not be stable: it can be revised as new information becomes available, and as the individual forming it changes own beliefs, desires, intentions, and evaluations.

That a sign can refer to multiple objects and/or concepts begs the question of where the limits to that multitude lie? Events and characters in Herman Melville's *Moby-Dick* have been taken as signs for many very different objects, acting as a hunting-ground for the symbolic critic, as Elmer Stoll had put it in the 1950s (Stoll 1951). The conflict between Captain Ahab and the white whale refers, according to some, to the battle between the consciousness of the white race and of its own abstract intellect, with the latter hunting to destroy the former. Others have seen

the whale as the Jungian (personal or collective – it is not very clear, but neither is Jung) unconscious, though that is certainly not a unique far-fetched reference. More recent discussions involve, e.g., the question of whether the post-September 11th USA can read *Moby-Dick* as anything but a story of revenge (Donoghue 2003). Perhaps, Melville intended the novel as a trick on those who venture far to hunt for the referent.

A less benign case of unexpected reference is a reading of war in the Italian Futurism movement. Initiated formally through Filippo Tommaso Marinetti's *The Founding and Manifesto of Futurism* published in 1909 in *Le Figaro*, it advocated breaking away from the past, abandon of intellectual and cultural heritage, and praise for danger, audacity, and revolt. One of its tenets, influential later in Fascism, was the following understanding of war:

[W]e Futurists have rebelled against the branding of war as anti-aesthetic [...] Accordingly we state: [...] War is beautiful because it establishes man's dominion over the subjugated machinery by means of gas masks, terrifying megaphones, flame throwers, and small tanks. War is beautiful because it initiates the dreamt-of metalization of the human body. War is beautiful because it enriches a flowering meadow with the fiery orchids of machine guns. War is beautiful because it combines the gunfire, the cannonades, the cease-fire, the scents, and the stench of putrefaction into a symphony. War is beautiful because it creates new architecture, like that of the big tanks, the geometrical formation flights, the smoke spirals from burning villages, and many others [...] Poets and artists of Futurism! [...] remember these principles of an aesthetics of war so that your struggle for a new literature and a new graphic art ... may be illumined by them! (cited in Benjamin 1999)

Similar perspective is on occasions depicted in Francis Ford Coppola's 1979 war drama *Apocalypse Now*. In a memorable scene many military helicopters fly toward the camera, set against the setting sun on a jungle horizon. In another bomber planes drop napalm on the jungle, fire and smoke rise. As Walter Benjamin, a literary critic, observed, there is an arresting clarity in Marinetti's way to assign unexpected properties to war. Coppola, on the other hand, has Colonel Kurtz, played by Marlon Brando, to negate in his last words "the horror...the horror..." whatever beauty Marinetti may have imagined. That there is clarity in a reference has nothing to do with it being desirable.

How wide ranging the reference will be and how unlikely the target of the original sign both depend crucially on the construction of the argument that supports the claimed reference, and ultimately on the willingness of the audience, of the recipients to accept the argument. Jungians may accept Melville's white whale as the unconscious for it may be presented as fitting within the worldview that Carl Jung volunteered. In contrast, a behavioral psychologist would hardly follow suit. Arguments need not appeal to evidence or manifest rigor. The symbolism of color for instance arises out of folklore, and is accepted because it is part of tradition, and tradition for many remains unquestionable. It is not a matter of reason to associate black specifically with mourning, as in the Judeo-Christian tradition, or to relate – as Shakespeare did in *The Winter's Tale* – yellow to jealousy (Heather 1948). If others did it before, and if no serious objections stand today for continuing in their path, then the path may well be good enough.



## ***Metaphor as Intentional Mis-Reference***

It would be an error to think that unexpected references are confined to arts or folklore. Management science has referred organizations on various occasions to the concepts of machine, organism, brains, cultures, political systems, and psychic prisons even (e.g., Morgan 1980; Keys 1991). To do so is to use metaphor as a tool for creative thinking. Organizations, we are told, are means – machines – achieving premeditated ends, operating ideally just as well-oiled and actual machines do. Once this reference is in place, we are led to claim that organizations should be efficient, predictable, and robust, just as actual machines are. To think of an organization in such terms of course poses problems, for it says that the machine must be re-engineered when its operating conditions change, along with its inputs. Hence the reference to an organism: once that reference is accepted, terms such as adaptability and evolution can be voiced – the organization adapts, some die out because they fail to adapt. Referring the organization to a brain is supposed to highlight that information plays a role in decision making, and that action follows choices made as the result of that information processing. Now, there is not much place within a machine, or a biological organism for, say, esthetics or ethics of its parts. It was thus to be expected that the organization would at some point start to be referred to as a culture, so as to highlight the role that specific esthetic, ethic, and moral guidelines, among others, can play in the activities of the organization's members. Then, there are office politics, the power plays within the organization that needed to be studied, which went together with the reference of organization to political systems, in which conflicts of interest are resolved through the use of informal authority. A final, somewhat odd reference we will mention is that of psychic prison. While not an outright mental asylum, an organization may apparently share the properties with collectives, sects perhaps, in which members take a number of assertions at face value. These form a background, against which any reference is made: adopted as beliefs, desires, intentions, and evaluations, they distort any information that the individual subsequently acquires. Perhaps a firm has provoked an ecological disaster, but any pointing out of such an event will then be seen by its members as an attempt to discredit what is otherwise – in their eyes of course – an organization valuable to its community. Having considered the machines, living organisms, as the referents of an organization, the least we can do at this point is observe that all of these unexpected references certainly do provide an eclectic and rich set of properties of organizations. A comprehensive intensional definition of the term (*an*) *organization* would clearly be quite complicated if it has the ambition to cover them all.

The method applied above in describing and prescribing the properties of organizations is simple. Take some sign, say  $S$ , and two referents, let these be the objects  $O_U$  and  $O_E$ , and call the former the *unexpected* referent, and the latter the *expected* referent. Put then a reference relation in between  $S$  and  $O_U$ , and another from  $S$  to  $O_E$ . If  $O_U$  is a particular, consider one by one the properties of the concept that it instantiates. For each property, discuss whether  $O_E$  instantiates that property as well if  $O_E$  is a particular; if  $O_E$  is a universal, then discuss whether  $O_E$  can incorporate that property. In case  $O_U$  is a universal, then consider each of the properties



it incorporates. For each property, discuss whether  $O_E$  instantiates that property as well if  $O_E$  is a particular; if  $O_E$  is a universal, then discuss whether  $O_E$  can incorporate that property. What is effectively happening here is a two-step process. First, a sign is referred to an unexpected referent (e.g., the organization is referred to a living organism) and to an expected referent (e.g., the organization is referred to the concept of a corporation, which groups the properties assigned to it via commercial law and other legislation). The second step is to consider one at a time the properties of the unexpected referent, and see whether the expected referent could share these properties, or could instantiate them. Of course, legislation is unlikely to describe corporation as having an “ability to adapt” to the changing economic conditions, and so we need to look elsewhere before describing or prescribing that property. Since legislation does not seem to mind the prescription that organizations be adaptable, the unexpected referent here – a living organism – provides a property that a particular organization may instantiate.

The question we should now turn to is the following: Is the management scientist and/or practitioner – when taking seriously these, so to speak intentional mis-references that seem to come a dime a dozen – not seeing the organization for what it is not? Is that individual not alike that other, who in Captain Ahab and the white whale’s conflict sees “the battle between the consciousness of the white race and of its own abstract intellect”? We should be careful in answering this. Stuck between the relative simplicity of the actually experienced organizational life – many do manage to perform well enough within organizations even though that admittedly involves work, ethics, office politics – and the elusiveness of an elegant and sound description thereof, both the researcher and the practitioner may well deem it necessary to turn to metaphors and allegories. Such a turn seeks another view on a subject of inquiry, for it leads one to consider properties that are typically not seen in the referent. Where the critique of the turn has its proper place is when there is a belief that the turn itself, the properties it points to do indeed describe the qualities of various particular and actual organizations. It is a mistake to describe the expected referent  $O_E$  by seeing it as instantiating the properties of the unexpected referent  $O_U$ . The unexpected referent cannot shape the expected referent; an organization is quite obviously *none of* machine, organism, brain, culture, political system, and psychic prison. The metaphor points to potential properties and does – should do – only that. These in turn are a starting point for further inquiry. If a living organism adapts to its environment in some way, we can relevantly ask whether an organization changes as when its environment does. But it should be at most a figure of speech when it is said that some effort in the organization is:

...involved in the comparison of internal structure and processes with external activity and trends and acts as it were part of the nervous and sensory system in the organization. [...] Organisms perceive changes in the environment and react as quickly as is possible in ways which maintain their likelihood of survival. Organizations [...] act similarly in attempts to remain viable into the future. (Keys 1991, p. 438)

Troubles arise when the individual hearing or reading the above has enough contextual cues to assume that its author says this not merely as just another way to more intricately say “members of the organization analyze the conditions inside

and outside of the organization in order to adjust internal procedures and interactions with external partners, clients, and others, in order to continue working as well as before, or better.” If there is something more to take away from the quote than that, then the use of the metaphor leads the recipient to mistakenly form reference. Metaphor requires the use of signs that are not typical for the object of interest: the terms organism, nervous system, sensory system, and survival all already have charged references, that is, often specialized references in particular fields. They certainly have proper definitions in biology. We are left wondering then how much of their established references of their original definitions does one know or wishes to carry over when using them in a different context.

With regards to our earlier question, many certainly will fall prey and see more than there is. It is not cynical to conclude that the trouble with metaphor, as a case of open reference, is that it requires exceptional care both when used in communication, and this on both the giving and receiving end. By requiring such care, it is bound to fail: worse than using signs local to a context, the metaphor will introduce those outside of it, and thereby deprive the latter of the contextual cues of their prior setting, while mixing them with contextual cues of the new one. Acceptable as a way to jump-start discussions on difficult topics, metaphor will prove problematic when some – otherwise achievable – rigor in reference is sought. The biologist who reads that an organization has the properties of a living organism will certainly form different reference for the term *organization* than will an organization scientist, who is not fluent in biology. Instead of using metaphor, the organizational scientist and practitioner ought rather develop own terminology and provide as rigorous as feasible definitions. Engineer an ontology in which there is no place for metaphor.

## 4.2 Vague Advice

Just how far can a person stand from the highest top of Mount Everest, and still say that she *is on Mount Everest*? Just how profitable must a firm be to be called *profitable*? How intensive should some resource be used for that use to be called *sustainable*? In all three cases, the question is essentially the following: What are the criteria that must be satisfied, before we can claim that some specific object (above, *a person, firm, use of resources*) instantiates some property of interest (*being on Mount Everest, being profitable, being sustainable*)?

When we lack definitive criteria to decide whether an object instantiates a property in interest, we face the problem of vagueness. We are unsure whether we can use a sign that refers to the property when describing the object of interest, for another person may have criteria different than ours. Communication that suffers from vagueness makes it difficult for the recipients to identify the referents. If we are interested in data on all profitable firms listed on some stock exchange, what specific firms are we interested in? Where is the cutoff point – what proportion of revenue should remain as profit for the firm to be called profitable? The crux of the problem is that there is a range of candidate referents for a vague sign, all of which

are all equally admissible as referents. To ambitiously venture to establish once and for all what precise referents are targeted by, say, the sign *tall person* will end with the recognition that there is no universal standard that would say what persons are tall, as opposed to those that are not. Looking at this in another way, from the object (e.g., a particular person) to the sign (e.g., *tall person*), we see that there is an uncertainty as to whether a sign (e.g., *tall*) can be used to refer to the object of interest (i.e., the particular person).

Vagueness is pervasive. Adjectives that have a comparative and superlative formations (e.g., tall, taller, tallest) are gradable adjectives and are a prominent source of vagueness. We spoke of essential properties in the third chapter; when the set of essential properties for a concept is not definite, then the sign used to refer to that concept is vague. Nouns can consequently be vague. Whether something is a chair is quite a question, because it asks what a prototypical chair is, and consequently, whether there is, so to speak, some culture-independent set of essential properties of chairs. Similar problem presents itself for verbs: when do we say that a person is running, and not jogging, or vice versa? According to Chris Barker, a philosopher, propositions (above, below) are not spared either.

A sign is vague if all of the following verify (e.g., Kennedy 2007 and related):

1. *Truth conditional variability*. Whether we would accept some sentence, which uses the sign of interest as true depends on the context in which it is used. *Tall, big, fast* are all relative. A bicycle moving at 40 km/h is a fast-moving bicycle, but an automobile moving at that same speed on a motorway is not a fast-moving automobile.
2. *Existence of borderline cases*. Whatever the context of use, there will generally be three sets of referents for a vague sign. Any object to which the sign refers without controversy will be in one of the sets. Another set will have all referents to which the sign clearly does not refer. The third set will contain the so-called borderline cases, that is, objects for which it is difficult or impossible to determine whether the said sign refers to. Taken out of context, the person of any height is a borderline case for the sign *tall person*. Similarly, any organization may be referred to as an *efficient organization*. If all chocolate bars in a shop have prices between \$1 and \$20, then one costing \$1 is cheap, another costing \$20 is expensive; is a third that costs \$10 expensive or cheap?
3. *The Sorites Paradox*. When employed within a particular form of argument, the sign will give rise to the Sorites paradox. The name is derived from the greek *soros*, translated as heap. The argument is usually laid out as follows (e.g., Hyde 2008):

1 grain of wheat does not make a heap.

If 1 grain of wheat does not make a heap then 2 grains of wheat do not.

If 2 grains of wheat do not make a heap then 3 grains do not.

...

If 9,999 grains of wheat do not make a heap then 10.000 do not.

10.000 grains of wheat do not make a heap.

The argument above appears valid, premises true, yet the conclusion false. While the argument proceeds by addition, it can be reformulated to proceed by subtraction: start by claiming that 10,000 grains of wheat do make a heap, and subtract grains of wheat to arrive at the undesired conclusion that 1 grain of wheat does make a heap. Same argument, be it by addition or subtraction, can be often constructed. If we claim for instance that the highest point of Mount Everest is on Mount Everest, and then accept that any point at one meter on each side of the highest point is on Mount Everest, we can proceed by adding meters to absurdly arrive at the conclusion that the entire planet is on Mount Everest.

Admitting vagueness in advice can leave the recipient puzzled as to what the advice is referring to. The problem here is not that of going too far in the search for the referent, as is the case in metaphor for instance. It is rather that there are potentially many plausible referents, but we lack the criteria to know which of them are being referred to via the vague sign. Vague advice delays commitment to a particular course of action: instead of recommending precisely what to do, vague advice may recommend whole ranges of options, and thereby leave the choice open. When taken too far, vague advice becomes no advice. To advise an individual to concede in a standoff begs the question of how much to give away before he has conceded enough.

It is a platitude that diplomats subtly use language, always conscientiously and deliberately choosing words. Vagueness keeps a privileged place in the diplomat's toolset precisely because it leaves open the debate on what the vague sign is exactly referring to. The multilateral trade negotiations, known as the Uruguay Round that lasted from the mid-1980s to 1994, ended with the formation of the World Trade Organization (WTO). Because they concerned sensitive questions of, e.g., import and export policies and intellectual property, a phrase – “special and differential treatment” – was intentionally introduced to avoid asking developing countries to commit to the same rules as other participants in the negotiations. What treatment is special and differential remained to be negotiated subsequently. The introduction of the vague term thus allows the conceding side in negotiations to delay commitment. This practice is hardly uncommon; the phrase *multifunctionality* of agriculture from the terminology of the WTO refers to benefits that agricultural policies can provide. It has proved most useful in defending protectionist policies before the WTO (Scott 2001). Since the term aims to refer to a wide range of benefits, called nontrade benefits (e.g., employment, protection of biodiversity, sustaining agricultural tradition), it is defined as a vague term. Moreover, since the list of its referents is not closed, there remains a possibility to build an argument that supports yet another referent, and in doing so advance own aims.

The origin of vagueness often is the presence of a predicate headed by a gradable adjective, e.g., *being tall*. Such a predicate designates a property of having a degree of height that is at least as great as some standard of comparison of height, that itself is not part of the usual reference for *tall*, but is determined by the context in which the said adjective is used. It will then depend on the context whether we will accept the use of that predicate to refer to some objects, or in other words, whether our use of that predicate will allow the audience to form reference, to know what

objects we are referring to. An adjective is gradable and can be assumed to give rise to vagueness within the sentence in which it appears, if the following conditions are met (Kennedy 2007):

1. The adjective maps its arguments onto abstract representations of measurement, or *degrees*.
2. The set of degrees totally ordered with respect to some *dimension* (e.g., cost, size, etc.) constitute a *scale*.
3. The adjective itself does not incorporate a standard for comparison, so that such a standard varies with context.
4. It should be possible to identify borderline cases of application of the given adjective.
5. When the standard for comparison changes, the sentence that includes the adjective may no longer be acceptable as true.
6. The property referred to by the adjective can be used in lines of reasoning that follow the one taken in the Sorites paradox (see above).

The first two conditions above assume that there is an ontology with the terms *degree*, *dimension*, and *scale*. The idea here, developed by Christopher Kennedy, a linguist is to start from the intuition that a gradable adjective comes together with its opposite, e.g., *tall* with *short*, *big* with *small*, *efficient* with *inefficient*, *intelligent* with *stupid*. The reason we see them as bundled together is that they provide “the same kind of information about the degree to which an object possesses some gradable property (for example, both *tall* and *short* provide information about an object’s height), but they do so from complementary perspectives. The positive adjective *tall* is used either neutrally or to highlight the height an object has, while the negative adjective *short* is used to highlight the height an object does not have” (Kennedy 1999, p. 75). For a given gradable adjective, we thus define a *scale* as a linearly ordered set of points, while a *dimension* indicates the property, the instances of which are referred to by the points on the scale. Hence, one specific point on the scale refers to an instance of a property. To say then that the gradable adjective maps its arguments onto degrees is to say that a gradable adjective relates objects (i.e., *person* in *tall person*) that it describes to degrees: a gradable adjective is consequently seen as a function from objects to degrees on a scale. The degree on a scale is simply a convex and nonempty subset of the scale; that is, a subset of the scale that satisfies the following property: for any two distinct points  $p_1$  and  $p_2$  on the scale  $S$ , and any third point  $p_3$  on  $S$ , if  $p_1$  and  $p_2$  are in the degree  $d$  on  $S$  and  $p_3$  is between  $p_1$  and  $p_2$ , then  $p_3$  is also in the degree  $d$ .

Given the scale, dimension, and degree concepts, a pair of gradable adjectives splits the scale into two parts, one referred to by the positive adjective, the other with the negative. To say *tall person* thus maps the specific person to the positive degree on the scale having *height* as its dimension. To say *short person* maps the person that is being referred to onto the negative degree of the scale for height. To define the standard of comparison is to choose a point on the scale which splits the scale onto the positive and negative degrees.

Although the analysis above applies specifically to gradable adjectives, similar reasoning can inform the analysis of other grammatical elements that engender vagueness. Consider the problem of evaluating whether some specific individual we are looking at is running or not. Even if we picture a scale that we would partition onto running and not running, the dimension we could associate with it seems rather different than, say, height used in evaluating if someone is tall or short. An important difference is that evaluating whether someone runs or not involves taking into consideration several properties, while height seemingly involves the observation of a single one. The question that should be asked is what the essential properties of running are, and which of these are satisfied by the individual that we are observing. Perhaps an essential property is that the individual is moving fast, while another is that her knees move higher than when he is walking. In both cases, and from *moving fast* and *move higher*, we can do the same as we did for *tall* and *short*. Namely, we can have one scale for *speed*, another for *vertical knee position*, then decide on the comparison standard, or in other words decide, respectively, on what speed is fast, as opposed to slow, and what vertical knee position is high, as opposed to low.

To understand vagueness via scales, dimensions, and degrees begs the difficult question of mental representation: what is the form of mental representation? Namely, when people think, e.g., to solve a problem, in what way do they manipulate the information available to them and what form that information takes within their minds. When wondering about form, we ask whether it is some symbol that somehow appears in the mind, or something else. Do people think of vague properties and their applicability via the ontology for the analysis of vagueness, that is, scales, dimensions, and degrees? While there is no definite answer, attractive arguments have been offered toward that line of thought. Peter Gärdenfors, a philosopher, suggests that information is interpreted via positions in structures called conceptual spaces (Gärdenfors 1996). As a simple example, consider the information acquired via those organs, which together enable the sense of taste. If the human perception of taste is generated from four distinct types of receptors – say, for saltiness, sourness, sweetness, and bitterness – then the conceptual space for taste is four-dimensional: if an individual says that two foods are of a similar taste, then these are close of the four dimensions of taste, that is, are on nearby positions in the conceptual taste for taste. The theory of conceptual spaces needs the same assumptions and ontology as the approach to the analysis of vagueness:

1. People can perceive or otherwise experience qualities of objects.
2. To experience a quality is to position, so to speak the result of that experience, called *quale* (plural: *qualia*), on a quality dimension.
3. A quality dimension is associated with a topological or metric structure.
4. A conceptual space can consist of one or more quality dimensions.

Now, there are two important clarifications to be made here, first on how conceptual spaces fit into our earlier discussion of vagueness, and then on the notions of *quality*, *quale*, *quality dimension*, and *conceptual space*.

To fit conceptual spaces with the treatment of vagueness, we should note that those scales we partitioned earlier onto degrees in dealing with vagueness are examples of topological structures, having the properties that all points are linearly ordered. Hence, a scale is one kind of topological structure that can be associated with a dimension. Consider the notion of time: if we call the dimension *time*, one plausible topological structure for this dimension is the line of real numbers. If this topological structure is chosen, then the *time* dimension is associated with a linear scale. In terms of the *scale*, *dimension*, and *degree* ontology, the term *dimension* corresponds to *quality dimension* in the ontology of conceptual spaces, *scale* to one kind of *topological structure*, and *degree* to a convex set of *qualia*. With conceptual spaces, we have basically the same notions, except for the (linear) *scale*, which is a special case of a *topological structure*.

We mentioned toward the end of the third chapter, when discussing what advice is not, the term *quality*: we said that the term refers to anything we can perceive or measure on any object, such as color, shape, size, weight, and which are inherent to objects they are perceived, measured, or otherwise experienced on. A quality of a particular is itself a particular: the shape of this tree is a quality specific to this tree alone, and inheres in the tree – that specific quality cannot exist without the tree. A *quale* is the result of observing or otherwise experiencing a quality of an object. It is a point on a *quality dimension*. Consider the perception of color as an example:

The subjective sensations called color are generally portrayed as a series of planes organized according to hue, saturation and brightness: hue is the quality of a color (i.e. its relative redness, greenness, blueness or yellowness); saturation is the degree to which the color differs from neutral gray; and brightness (or lightness) is the perceived intensity of the color [...]. Any one of these planes comprises four primary color categories (red, green, blue and yellow), each of which is characterized by a unique color percept (the approximate position of which is indicated by dots), that is, a color experience that cannot be seen or imagined as a mixture of any other colors. (Lotto and Purves 2002, p. 85)

If we follow this line of thinking, our experience of the red color of a rose gives three qualia, each on a different quality dimension. The three quality dimensions are hue, saturation, and brightness. Continuous linear scales are adequate as topological structures for saturation and brightness. The topological structure for hue can be pictured as a circular plane, on which each point is at some distance from each of the four primary color categories. A color quality of an object is experienced via three qualia, each taking a position on its quality dimension. The three quality dimensions together form the color space. A *conceptual space* is made out of one or more quality dimensions.

We can think of intensional definitions as referring to conceptual spaces. A written definition will use signs, each of which will refer to properties. When we say that this rose is red, we are saying that the rose we are looking at is instantiating the *red rose* concept. We would expect that concept to include a property, such as *being red*. Given the theory of conceptual spaces, we can define the property *being red* as a region in the conceptual space for color. If we accept that color is perceived via hue, saturation, and brightness quality dimensions, then *being red* can be defined



as referring to some part of the conceptual space constituted of these three quality dimensions. In more general terms, we can assert the following:

- A quality is a particular that inheres in some particular object. The quality does not exist without the object that it inheres in. For example, the red color of this specific rose, where the rose's red color is its quality, and the object is the rose the individual is looking at.
- A quale is our subjective experience of a quality of a particular object. For example, that individual's perception of the red of the rose he is looking at.
- A quale has a position within a conceptual space. For example, if the individual experiences colors through three quality dimensions – hue, saturation, and brightness – the quale that results from the individual's looking at that red rose has a position on each of the three quality dimensions, that is, has a position in the conceptual space formed by those three quality dimensions.
- The object instantiates a property if the quale – of the individual experiencing a quality of that object – falls within the region of the conceptual space, to which the sign of the property is referring to. For example, the individual is looking at that rose (object); the individual experiences the red color (quality) of that rose as a quale in her subjective conceptual space for color; if the quale falls within the region referred to by *being red*, then that individual sees the rose instantiating the property *being red*.

A property is a region in a conceptual space, so that an intensional definition refers to regions of conceptual spaces. This insight helps in designing advice: ideally, the properties advice refers to will be regions in conceptual spaces that both the giver and recipient of advice will share. In many cases, conceptual spaces will be learned from standards. If we speak of weights, sizes, and distances, among others, we are likely to share similar conceptual spaces. In the case of weights, maybe we both share the metric quality dimension for weight, and in case of a mismatch (e.g., one uses imperial, another metric) we can still agree relatively fast, as conversion conventions have been established and measurement instruments are available. It is, however, overly optimistic to expect that conceptual spaces are often shared. Despite the many cases of learning and standard conceptual spaces, many situations require us to refer to properties that are regions in, so to speak, not yet properly explored conceptual spaces. In such cases, a term may be vague not because of a missing definition, but because the definition refers to properties in conceptual spaces that are not shared.

An advisor who designs recommendations on, say, how to improve the quality of a product or service faces the same problem as Plato, who cannot know if a just man can be happy, for he does not know what justice is. Joseph Juran, a twentieth century management consultant who helped the introduction and development of quality management practices, offered one definition of quality as the extent to which a product or service corresponds to the purposes of its user (Juran 1951). While this certainly helped organizations understand that they should seriously consider what the users may be expecting of the products and services and not only what their engineers believe is adequate, it did not – and could not – say



much about what the users do expect. Consider carefully what happened in that turn to consumers' perception of quality in place of only the engineers' perception of quality. If an individual working on the design of a product or service is asked to define that product, he will refer to regions of quality spaces that he has learned or constructed throughout the time he has been designing it. It is not unexpected for this individual to be able to identify many properties of the product. The conceptual space, in which these properties delimit a region, may be entirely unlike the conceptual space, comparatively less elaborate, of an individual who is only occasionally using that product. To ask, consequently, the occasional user to explain why he may think that the product has, say, a good quality, will give answers that may be very different from the engineer. One explanation for this difference is that the engineer and the occasional user do not put the product to the same use. Perhaps that is a good enough explanation, but another one is certainly relevant: that the two individuals are interested in the product to different extents, have consequently learned about the product to different depth and, finally, use different properties in different conceptual spaces to describe the same product. It is trivial to see this, by comparing, say, the engineering specifications of products and any description that a customer may give: the former will involve as well defined as feasible quality dimensions and associated properties, since the absence of these two would render quality control impossible; the customer will, in contrast, provide a much less elaborate description. We thus get to the point: the turn to consumer's perception of quality has introduced a serious difficulty in establishing what precisely quality is for a customer, for it requires finding out about the customers' conceptual spaces. Research in marketing science has recognized this problem. In research on the quality of service provided to customers, it has been suggested that the customers evaluate quality through eight dimensions; as an example, consider two of them:

Reliability involves consistency of performance and dependability. It means that the firm performs the service right the first time. It also means that the firm honors its promises. Specifically, it involves: accuracy in billing; keeping records correctly; performing the service at the designated time.[...] Courtesy involves politeness, respect, consideration, and friendliness of contact personnel [...] It includes: consideration for the consumer's property [...]; clean and neat appearance of public contact personnel. (Parasuraman et al. 1985, p. 47)

Each dimension is defined both via properties and examples, so that each definition is a combination of an intensional and ostensive definition. The properties referred to in definitions are not straightforward. Dependability, respect, consideration, and friendliness are vague terms and refer to regions of subjective conceptual spaces, hence the need to provide examples of more concrete actions. In pointing out particular actions, it is hoped that they will produce effects, which will bear qualities, and finally, that the customer will experience these qualities as qualia in conceptual spaces referred to via such terms as dependability, respect, consideration, and friendliness. It is clear that the chain is long, starting in actions that a company performs and ending in the customer's experience of their effects. Perhaps, a gap is a more appropriate metaphor than a chain here: a gap from the intentions on the side of the firm, to the reading of the effects of these intentions on the

consumer side. This length of the chain, or the breadth of the gap, indicates the risk of mis-reference and gives advertising a purpose alongside that of explicit promotion: recipients need to be taught conceptual spaces. We would consequently expect the content of advertisements to define conceptual spaces, in which the consumer should inscribe her experiences of the product or service; but do they?

A full-page ad for an IBM personal computer, published on 26 October 1981 in *Time* magazine is split onto three parts: (1) text, carrying sentences, such as “Features like high resolution color graphics. Ten user-defined function keys. The kind of expandability that lets you add a printer for word processing, or user memory up to 256KB.”; (2) a table captioned “IBM personal computer specifications” with a detailed list of properties that its engineers use in describing the product; and (3) an image of a joyful businessman alongside the unit itself. Qualities of the computer are explicitly given and are suggestive of criteria that the consumer should account for in the purchase decision. Degrees are specified by suggesting that the memory of that particular personal computer is superior to those available at other brands.

Not all advertising, however, approaches the teaching of conceptual spaces in such a manner, especially when imagery dominates. To have the nontextual elements as prominent ingredients could seem surprising, given how open reference can be. Benetton, the Italian clothing manufacturer, is a case at hand; the company explains the signs used in its advertising campaigns as follows: “through innovative, sometimes provocative advertising campaigns, the Benetton name has become synonymous with multi-cultural diversity, inter-racial harmony, and an upscale approach to fashion retailing [...] different markets respond differently to different trends from the collection and are experiencing different climates and seasons at any particular time, therefore it is virtually impossible to represent our 4,000-piece yearly product offering” (cited in Leslie 1995, p. 414). The point here is that Benetton’s advertising is not trying to teach quality dimensions to the consumer, in the way IBM did with its personal computer; if so, Benetton would market the qualities of the product itself. Their solution is rather to present signs that refer to conceptual spaces unrelated to the products themselves, such as multi-cultural diversity. The context in which such references can be established is constructed by the media, art, and education. Instead of teaching the consumer conceptual spaces through which to experience the product, and ultimately compare it to others, Benetton is using signs that can be experienced within conceptual spaces that the consumer already uses. The beneficial business effect is that the approach scales exceptionally: whatever the number of different products sold, the same advertisements will do. The consumer is taught the reference relation between a particular sign – namely, the logotype or any sign that is repeated on the product, its packaging, its promotion – and the properties located within conceptual spaces that have nothing to do with the sign itself. This objective is achieved when individuals refer the sign to such abstract concepts as lifestyles, social or professional positions. Hence the term *brand equity* in marketing research and practice: “a brand is said to have positive (negative) customer-based brand equity if consumers react more (less) favorably to the product, price, promotion or the distribution of the brand than they do to the same marketing mix element [i.e., combination of product, price,

promotion, or distribution] when it is attributed to a fictitiously named or unnamed version of the product or service” (Keller 1993, p. 8). Experiments have shown that when individuals have been exposed to the signs of a brand before experiencing the actual product, the properties experienced from the brand signs will be more important in the individual’s evaluation of the product than the properties experienced from the product’s qualities (van Osselaer and Alba 2000). In our terminology, what is happening is that the substitution of conceptual spaces biases the experience that the individual has of the concrete qualities of the product. Not only can the individuals be taught reference relations from signs to properties, but once they have learned it, it can influence the learning of reference relations for other signs, as long as the former and latter signs are communicated together.

There are consequently at least three broad ways to, if not alleviate, then at least tone down the impact that vagueness can have in communication. One is to set a standard of comparison, to define degrees. For it to achieve its aim, to render more precise, the conceptual space needs to be shared – participants involved in communication need to know the scale and (quality) dimension within which the standard of comparison is set. When standards set the quality dimensions of a conceptual space, a unequivocal standard of comparison can be set. A second approach is to teach the scales and quality dimensions of the conceptual space to the recipient, then set a standard of comparison. A third way is the economical approach applied prominently in contemporary advertising. Potentially vague signs, or more generally, signs having open reference are communicated with others, the accompanying signs. How are the latter chosen? A candidate is a sign, or a collection thereof, for which the audience can be assumed to have already established reference relations. Whether some candidate is more appropriate than another will depend on the aim of the communication beyond the avoidance of vagueness. Hence, the careful choice of, say, whom a politician takes pictures with, and why their images with children are unavoidable in the language of political communication.

### 4.3 Referent Lost

Mis-reference seems plausible enough a phenomenon: contextual cues available in a communication prove insufficient for the recipient to identify the referent intended by the advisor; the recipient consequently forms the reference relation between the sign and some other referent. If it is believable that mis-reference can occur, is it then also believable that when it does, the intended referent must exist? Contrary to its appearance, this question need not be treated as a metaphysical one: we are not interested in whether any referent does in fact exist and of what kind it is. We can reformulate the question to make this clearer: When mis-reference does occur, can we *always* – through further communication – recalibrate the advised individual’s reference relation so that it points to the intended referent? Can we always make sure that the recipient ultimately does make the reference intended by the advisor?

To answer with a positive, we must make the following heroic assumption: whenever there is mis-reference, the advisor can bring the recipient to experience the intended referent and thereby establish a reference relation from the sign to the intended referent. There are many cases that violate this assumption, and consequently favor the negative answer. They all lead to this conclusion: in some cases of mis-reference, the advised individual cannot be induced to experience the intended referent through communication or otherwise. From the perspective of the recipient, the intended referent in such cases simply does not exist. A reference cannot be more open than when there is no referent.

In Stanley Kubrick's 1960 adventure drama *Spartacus*, when the rebel slave laments to his beloved Varinia that although he has finally freed himself, he still is ignorant, she asks him what he would want to know. He answers, "Why a star falls and a bird doesn't. Where the sun goes at night. Why the moon changes shape. I want to know where the wind comes from." She only answers the last of these: "The wind begins in a cave. Far to the north, a young god sleeps in that cave. He dreams of a girl. . . and he sighs. . . and the night wind stirs with his breath." Beyond its romantic character, the scene fits very well this fictional story of emancipation, read by many as a metaphor to the struggle of the humble, ignorant but interested, and poor proletariat, with its decadent and wealthy oppressor. Now, suppose that the story takes a different turn after that scene. Instead of pursuing with warfare, Spartacus takes none of her explanation for granted. Suppose that, interested by her explanation, his desire to know overwhelms his will to fight, and that, out of respect for Varinia's love and for lack of a better explanation, he sets out to find the wind god. If we remain in fiction, we could very well have him find something far to the north, as she instructed. Our skeptical modern view is unforgiving: if he was to go north, manage to explore many caves, he would nevertheless fail to reach the end of the world as a geographer of the time would expect him to. Unable to make sure that he did explore enough caves, perhaps he would live long enough to go so far north to find himself south of the starting point, thereby failing to find a northern cave. A contemporary skeptic would advise Spartacus against such a quest; there is no god of wind in some cave far to the north, he would argue, it is only a myth.

### *Antireferent*

Myth and taboo incorporate and perpetuate signs with open references. If Spartacus has no way whatsoever to experience the same referent as Varinia, when she speaks of the wind god, then her explanation has no referent for him. This conclusion remains standing even if they share the same faith, myths, and taboos, and thus accept that there is something they call the wind god, and that it does lie in a cave, longing for a girl. The whole point of putting that wind god in a faraway cave is to mark its inaccessibility, hence two positions. If we accept from the outset that such a thing does not exist, then any experience thereof that an individual may claim will be an experience of something else, mis-referenced to the *wind god* sign. The

second position is this: if we claim at the outset that it does exist, but still remains inaccessible to experience, and not amenable to agreement on its existence, then what remains justifiable is that there is indeed no referent.

We should be careful here, to avoid thinking that, if the referent was accessible, the experience thereof may well be amenable to description, which in turn could be validated or invalidated by those sharing the described experience. They may well do the validation, but this does not mean that they do share the same conceptual spaces and that they do experience the referent in the exact same way (contrast, e.g., how daltonists perceive color, as opposed to those who do not share that problem). Rather, it would merely mean that *they have agreed that their remaining disagreements are tolerable for their practical purposes, for which they must coordinate.*

It is *prima facie* evidence that experiences are personal, or subjective. In being such, an individual can neither know the specifics of someone else's experience nor the other's conceptual spaces: there is neither mind reading nor telepathy. The skeptic would then undoubtedly ask this: how can then we speak of *shared* conceptual spaces? Or, in terms closer to the sorites paradox: How much should people know about each other's conceptual spaces so that we can say that they share these conceptual spaces? There is effectively the sorites paradox here, but to focus on it is to miss the point altogether. It is also *prima facie* evidence that the satisfaction of personal desires requires collective effort, demanding thus coordination, and coordination in turn cannot happen without communication. Even in the case of a sociopath murderer who kills alone, coordination is not absent: while he may not need to coordinate to satisfy the urge, others will coordinate to stop her from doing so. When Gene Sharp speaks of nonviolent action, he does not speak of the absence of coordination toward violence, but of the presence of coordination toward nonviolence. When the community accepts some individual's behavior, its acceptance is a reflection of its coordination to accept, not of its noncoordination to reject: the dictator's power comes from the eyes of its subjects, not from some source external to them all; when subjects do accept the dictator's authority, they do coordinate in their acceptance, for it is their acceptance of the meta-advice in the political system that translates into coordination. Returning to conceptual spaces and antireferents, the skeptic does point relevantly her finger to vagueness, but this in no way affects the fact that, if desires are to be satisfied, coordination cannot be postponed. The cliché applies: one does – must do – either with what one has, or can design. It is not this obvious observation that matters much, but that one must also know the problems hidden behind the cliché. It is to understand that signs referring to antireferents are tools of communication much more versatile than signs the referents of which can be experienced, such as, say, this apple, this building, or this desk. It is within the grasp of most to experience a particular apple, building, desk. Is it so also for notions such as globalization and rationality? Can one experience “the process of globalization” in the same way as one experiences a particular apple he is looking at, or eating?

Many have gone out of their way to explain what it is to be rational, especially so that we can know what is irrational and somehow undesirable. Not unlike the commandments in the Christian tradition, economics has its axioms of perfectly rational

economic behavior. If the individual fails with regards to the commandments, he will not be a good Christian, or so the story goes; in mainstream economics, if the individual fails to behave in line with the axioms of rationality, he is irrational. The standard axioms of perfectly rational choice come from the works of John von Neumann and Oskar Morgenstern (von Neumann and Morgenstern 1956), Leonard Savage (Savage 1954), and Howard Raiffa and Ralph Keeney (Raiffa 1968; Keeney and Raiffa 1976); one formulation thereof is below (Keeney 1982, pp. 830–831):

- The individual can identify at least two alternative courses of action.
- He can identify the possible consequences of each alternative.
- The relative likelihoods (i.e., probabilities) of each possible consequence that could result from each alternative are also known.
- The relative desirability (i.e., utility) for all the possible consequences of any alternative are known.
- If two alternatives would each result in the same two possible consequences, the alternative yielding the higher chance of the preferred consequence is preferred.
- If one alternative is preferred to a second alternative and if the second alternative is preferred to a third alternative, then the first alternative is preferred to the third alternative. That is, preferences are transitive.
- If an alternative is modified by replacing one of its consequences with a set of consequences and associated probabilities that is indifferent to the consequences being replaced, then the original and the modified alternatives should be indifferent.

At least since Herbert Simon, a twentieth century economist, management and computer scientists, and psychologist formulated his arguments for a bounded and not perfect rationality in the 1950s (Simon 1955), the axioms above have been attacked so much that they should mostly remain of historical interest, yet they still keep a prominent place in economic theory. They tell us that when facing a choice, the perfectly rational economic man identifies if not all, then impressively many possible alternative courses of action. Alternatives are subsequently ranked according to a well-organized and stable system of preferences. The alternative is chosen that dominates all others on the decision-maker's preference scale. Is this what actually happens, or in other words, do/can we experience such behavior when we choose? Apart from *prima facie* evidence, many elaborate arguments have been given to favor the negative answer, including Daniel Kahneman and Amos Tversky's empirical work in psychology (Kahneman and Tversky 1979; Tversky and Kahneman 1981). If perfectly rational behavior is just as inaccessible to the individual as the wind god to Spartacus, is the referent of perfectly rational behavior not an antireferent? In other words, is it not a referent that is elusive to experience, and hence, a designed referent that can only be understood in terms of its purpose?

While any sign is a tool of communication, some signs do have the desirable property that their intended referent is accessible to the recipient, who can thus experience it under some reasonable conditions and guidance from the advisor. For such signs, it may be possible to reduce mis-reference through further communication with the recipient. Other signs refer to antireferents, that is, their referents

are not amenable to immediate experience, or could have been the subject of an individual's prior experiences. An inaccessible referent is as good as an inexistent referent. As it escapes experience, it is an antireferent, admitting any intensional or ostensive definition. Any behavior can be rational, as the definition of rationality hangs entirely on what we may least disagree on.

In contrast to a sign having an accessible referent, a sign with an antireferent carries a twist: its antireferent can be reified, that is, an individual may paradoxically end up believing that the antireferent does exist. It is, for instance, to believe that there is something called perfectly rational behavior, and perhaps alongside, to believe that we would all be better off if people would just learn to choose better, to get closer to that decision behavior. It is to believe that there are conceptual spaces, that people *do actually think* in terms of quality dimensions, that objects *do actually have* qualities. Any antireferent dies is lost when belief in it fades. There is perfectly rational behavior only in the worldview of the individual believing in it, it is only accessible to him.

The term *the People* used in communist regimes refers to an antireferent: the Party believes that its authority comes out of the People's will for the Party to rule, because the Party promotes the will of the People. The role of the referent for *the People* to legitimize the rule of the Party is not unlike that of *the God* to legitimize the rule of *the King*: in both cases, it is the ruler, respectively, the Party and the King, that will define the antireferent of, respectively, the People and the God. As Slavoj Žižek, a philosopher observes:

the Party thinks that it is the Party because it represents the People's real interests, because it is rooted in the People, expressing their will; but in reality the People are the People because — or, more precisely, in so far as — they are embodied in the Party. [...] The paradoxical functioning of the 'People' in the totalitarian universe can be most easily detected through the analysis of phrases like 'the whole People supports the Party'. This proposition cannot be falsified because behind the form of an observation of a fact, we have a circular definition of the People [...] the real member of the People is only he who supports the rule of the Party: those who work against its rule are automatically excluded from the People; they become the 'enemies of the People'. (Žižek 1989, pp. 164–165)

Just as it is the Party, and not the citizens that designs the referent of the term *the People*, so it is not the mob ruled by the King, but the King that designs the referent of the term *the God*. The referent of *the People* does not exist, because this term is not introduced to put a name on some object accessible to experience. Rather, there is a need for the object, and thus a sign is introduced, and the object defined to suit the purposes of its designer. The term *the People* is designed before its referent, the referent begins to exist as soon as its properties are defined by the Party. Same observation applies for the term *perfectly rational behavior*, which is a name for a set of properties lumped together toward a particular aim, which most likely was the construction of relatively simple mathematical models of decision-making as a starting point for further criticism and research on choice behaviors.

There is a temptation to see mischievous intentions behind the use of terms having antireferents. The arbitrary accusation of being an Enemy of the People is performed through the use of the term *the People*. To uncover its arbitrary referent would not necessarily alleviate accusations, but it would lead to the conclusion



that the accused is not the Enemy of the People because he is the enemy of, say, each and every person. Rather, he is accused for some other reason, which remains unknown in the accusation itself. The designer of the term *the People* thus uses the term to intentionally induce the recipient, the hearer to mis-reference. If the King does as his God commands, there is no need for him to explain his acts, they are explained through the mis-reference to the inaccessible God. If the Party accuses because it follows the will of the People, then any further explanation of why the accusation is made becomes irrelevant: accusation comes out of the will of the People. The antireferent of the term explains of course nothing, since it is inaccessible to experience in the first place. It is its inaccessibility that makes it a versatile tool in communication. Its use leaves the recipient in the dark and lets the advisor illuminate the way. Facing ignorance, the recipient may succumb to thinking that although he cannot experience the referent, that there is something to it, that it is neither empty nor arbitrary. Belief in the good intentions and/or expertise of the advisor lets the latter impose mis-reference on the recipient.

It is an understatement to say that antireferents of many terms have been designed to disastrous consequences. It is enough to recall that *the People* was in recent history taught quite a bit about *Political Correctness*, *the Weapons of Mass Destruction in Iraq*, *the Axis of Evil*, *the Democracy in X*, where you can replace X with the name of some country that has had clear-cut totalitarian regimes for much of its past few centuries; Iraq and Afghanistan immediately coming to mind. Older such terms remain potent, and if you need examples, pick any name of a deity, regardless of faith. The Norwegian Nobel Committee announced that a reason for awarding Barack Obama, the 44th president of the USA, the 2009 Nobel Peace Prize is that “[h]is diplomacy is founded in the concept that those who are to lead the world must do so on the basis of values and attitudes that are shared by the majority of the world’s population” (The Norwegian Nobel Committee 2009). A laudable concept indeed, but many will remain wondering who has privileged access to those values and attitudes shared by the majority of the world’s population, who is capable of knowing them, and thus, who will design the antireferent of the phrase *values and attitudes shared by the majority of the world’s population*.

It is hardly controversial to accept, as Aristotle claimed (Crespo 2008), that shared values are a precondition to the predictability of individuals engaged in exchange. Myth and taboo seem to aim at uniform values, toward better coordination. This begs the question of whether reference relations arising from myths should be more or less open, for the myth to accomplish its coordinational role. Can there be coordination if the referent itself is missing? It might appear that coordination under myths requires that all individuals share the same reference relations, that they be unequivocal. However, the wind god case is a trivial illustration of a myth, in which the referent is inaccessible and, consequently, to some extent open. This is where we should remember that the more open (i.e., the less precise) the reference, the more freedom there is to ascribe arbitrary properties to the fictional/hypothetical referent. If Spartacus does accept that there is something, to which Varinia refers as the wind god, then Varinia can describe that something with whatever properties she desires. As long as she chooses properties that she assumes are believable



to Spartacus, she can himself define the referent. Her ability to do so comes precisely out of the referent's inaccessibility to Spartacus. In doing so, she teaches him conceptual spaces, within which to position the properties that she mentions (i.e., the sighing and the longing, the remoteness of the cave). For the Spartacus character in Kubrick's motion picture, the wind god is the obscure, inscrutable referent.

As the interests of the advisor change, so will the properties of the antireferent. The antireferent will take the shape fitting to the purposes of the moment. This is precisely why it is named *antireferent*: not only does it avoid a stable definition, it thereby also escapes a precise definition. Precision is effectively unnecessary as long as the properties attributed to the antireferent are sufficient for its designer to achieve the objectives she pursues. Properties and precision can be added along the way. Recall the term *multifunctionality* from the WTO; it is vague not because its designers are ignorant, but because it should be used to hit a moving target, which is to defend protectionist agricultural policies that contradict free trade rules. As such, there is no need to make explicit too many properties of its antireferent. It is moreover damaging to claim that some properties that are made explicit are also stable: different countries will at different times need protectionist policies, and all trading partners are well aware of that. No one thus has the interest to destroy the term by making it refer to a definite referent and not an antireferent.

A reference without the referent makes it possible to design the referent of one's own choosing. Myths illustrate that not many properties are needed – not much precision – for the sign and its designed antireferent to be relevant for coordination. It may be enough if it is an omniscient thing with an elaborate ability for observation and control, such as the deities central to various faiths. The mechanism for coordination is simple in such a case. First, the individual designs the antireferent as seems fit for whatever own or collective purposes; he does so by assigning properties to it, either undesirable so that it is feared, or desirable ones, so that it is admired. In the second step, best done in parallel with the first, the individual should proclaim a privileged access to the referent. He can consequently limit who it is that should be listened to, when the properties of the missing referent are sought. The overall aim of proceeding in this way is to put in place a mechanism of consent: if the referent is to be feared, then the fear of present or future consequences of disobedience facilitates consent; if the referent is admired, acting against it is to deny respect for its admirable properties. Meta-advice is dictatorial, asking one to unquestionably do as told.

It is too easy to denounce myths, taboos, and faiths as perpetrators of signs devoid of scrutable referents, in other words, of referents for which mis-reference can never fully be avoided. In the words of Ralph Waldo Emerson, a nineteenth century poet and philosopher: "The religion of one age is the literary entertainment of the next." Emerson's remark remains uncontroversial when paraphrased for advertising campaigns, party programmes, corporate mission statements, and ideals proclaimed in political systems. The ethical commitments of the infamous Enron corporation were already mentioned in the third chapter. In the context of their management's decision making, the phrase "highest professional and ethical standards" must have had a very open reference, to a point of either having no referent

whatsoever, or having a designed antireferent, with properties opposite to those commonly thought of as professional and ethical conduct.

When the referent remains open, that is, when its definition is not precise enough, it will be designed to suit specific aims. The properties of the referent should consequently be analyzed in light of the referent's designers' assumed objectives, and actions observed toward the realization of those objectives. Stated otherwise, advice is designed with a purpose, and the properties of the referents of advice are chosen to suit that purpose. With this in mind, the question for the advisor is how to design the referent – what essential properties to give to it – to achieve some aim. For the recipient, the task is rather different: how to establish if the referent is indeed designed, to what purpose, and consequently decide whether to accept or reject it. In less technical terms, the question for the recipient is how not to be a *homo follis*, an individual easily taken advantage of.



## Chapter 5

### Kinds of Advice

It is common to hear advice being qualified as good or bad. Advisors' reputation may depend on such qualifications, and consequently their current and future public's acceptance of the advice they dispense. To be good, advice should presumably lead the recipient to achieve the desired consequences in his decision situation. Advice could also be called good when the observed consequences of choice are undesirable, but the individual was advised against the chosen course of action. In the opposite cases, advice would be called bad: (1) the individual followed it, and the consequences turned out undesirable, or (2) the individual acted against advice, and the consequences turned desirable. All this appears straightforward, in particular, because the motive seems laudable: we hope to draw normative lessons about some advice and advisor after we have observed the effects of decisions. The causes for the collapse of the Enron corporation in 2001 continue to be intensely explored, for example. The motivation for the legal profession is that "[a]n examination of the confluence of ...[causes] provides insight into the type of reforms, regulatory or otherwise, that might prevent other unprecedented collapses or perhaps provide the means of intervention before the ...[causes] emerge." (Jennings 2003), or, in other words, "[t]he reasons for Enron's collapse should affect the normative conclusions of scholars..." (Partnoy 2003). It would be naïve to argue that qualifications of advice *must not* be made, as they clearly and often are. That they are made has nothing to do with whether the thinking that gives such evaluations is itself, so to speak, any good. Regardless of how rigorous the method applied to arrive at such qualifications, the problems that remain when choosing them should be understood.

Consequences of a choice can only be observed after the choice has been made. They are inherently neither good nor bad – whether they are desirable is in the eye of the beholder, for it is the individual experiencing these consequences who qualifies them. Consequences of decisions rarely receive a unique evaluation – important choices will have effects on more than a single individual. Economists speak of *externalities*: an economic transaction may be beneficial to all directly involved in it, but may affect others as well. Recall the case of fairs in twelfth century Europe, discussed in Chap. 2: a transaction should satisfy the principle of reciprocity, which states that (1) both parties should benefit from the transaction, (2) neither of them should bear disproportionate advantages or disadvantages compared to the other party, and (3) third parties should not be disadvantaged. Despite potentially good

intentions of the trading parties, this last condition can never be properly verified in practice. Negative or positive externalities to third parties may be wide ranging, and their effects striking only after a long period of time. But anticipating negative externalities is not only a matter of cost. As different people believe, desire, and intend different things, they may also differently evaluate the consequences of a choice. It follows that – because the good/bad qualification of advice depends on the good/bad qualification of the consequences of a decision – advice can also be evaluated differently by the individuals benefiting or suffering from the effects of the decision.

Subjective evaluation of the many consequences of choice is not the only problem that makes the evaluation of advice debatable. Another problem is the wrong assumption that causality can be uniquely established, or to borrow from Wittgenstein, that “[t]he events of the future cannot be inferred from those of the present. Superstition is the belief in the causal nexus” (Wittgenstein 2001, paragraph 5.1361). To follow the advice that proved “good” in the past, and expect desirable consequences precisely because the advice was so qualified, is but an instance of that superstition. A related problem is uncertainty, as even the most rigorous and best informed decision-making still cannot guarantee certain future success, for we cannot predict potentially consequential events.

With causality that is elusive, uncertainty that diminishes the relevance of prior experience, and success (or lack thereof) that is subjective, the qualification of advice in good/bad terms is objectionable despite it being frequent.

Any attempt to give general guidelines on how to recognize “good” or “bad” advice can thus happen only if we relinquish commonsense. Instead of speaking of good or bad advice, it remains possible to identify some characteristics that advice needs to have to be relevant to the recipient’s decision situation. We should, of course, ask when can a sign in advice be called *relevant*? To the extent that any advice is given to influence the recipient’s understanding of his choice situation – to affect his reflection on the choice problem and the final act of choice – we may conclude that advice is relevant if it has lead the recipient to do as advised. Matters are, however, more complex than that. The recipient may act as advised simply because advice coincides with the already preferred course of action. When the observed behavior deviates from advice, the cause need not be the inappropriate choice of signs. The recipient may perfectly well understand what the advisor was referring to and, of course, choose otherwise.

Individual autonomy and the openness of reference stand between the recipient and the observed choice. Misreference arises out of communication that uses signs with open reference, or in other words, the use of signs within a context that lacks some of the cues that the recipient needs to form the reference intended by the advisor. Independently of misreference, individual autonomy ensures that the individual need not blindly do as told, but reflect on what he has been advised and had concluded on his own about the decision situation.

Instead of calling relevant only the advice that coincides with observed choice, the very effect of an individual’s autonomy in choice needs to be left out of the discussion. What remains is, roughly speaking, that advice will be relevant if it is

understood by the recipient. This in turn begs the question of how to ensure that advice gets understood, or in other words, to ascertain that the recipient forms the references intended by the advisor.

Once we leave telepathy in its realm of fiction, actual communication with the recipient remains the only feasible way for the advisor to check whether the recommendations are relevant. This seems simple enough: if the form of communication and resources allow it, pursue the interaction to check, as much as possible, if the recipient formed the references intended by the advisor. Resources are limited in actual settings, and it will not be possible to come out of this verification with the certainty that references did form as intended.

Despite the unavoidable limits, what remains possible is a deeper understanding of (1) the form and content of information that any advice-giving/receiving communication may involve, and of (2) how these pieces of information may be combined with the information that the recipient uses in defining and weighing his options before choice. The first aim will be pursued already in part in this section. We will specialize the concept of advice, by building on the definition of advice offered in the third chapter and some standard theories of decision-making. The subsequent section will then offer an ontology of advice. Concepts in that ontology will in turn be combined within what we will call the *advisor's problem*, that is, the basic problem that any advisor needs to resolve when designing advice.

## 5.1 Classification from a Model of Choice

If there is little interest in classifying advice as either good or bad, what other broad categorization may be useful? We defined advice in terms of communication in Chap. 3. To have advice, we must have at least two people, the advisor and the recipient. The former performs speech acts, the dictum of which is made of signs referring to the decision context of the recipient. The aim of the recipient is thus to communicate to the recipient something about the decision situation the latter is facing.

Insofar as the purpose of advice-giving is to influence the recipient, it is necessary for the advisor to make assumptions about the recipient and his decision context. The advisor's other option is far less attractive, for it rests on a disregard for the presumed specifics of both the context of choice and of, say, beliefs and attitudes of the recipient. This other option is unlikely to result in relevant advice.

Relevant advice is produced on two sets of assumptions. One reflects the advisor's conception of the elements of the recipient's decision context. Take the case when advice comes out of the advisor's prior experience: the advisor may believe to have lived a similar choice situation to that of the recipient, and consequently deem useful to dispense advice. Before the advisor believes that he has already experienced the recipient's decision setting, he either needs to have discovered that decision setting himself, or in case it is inaccessible to him, must assume what may

form that decision context. Supposing that the recipient is to decide whether to purchase some specific automobile, the advisor can experience the decision setting if he is also present with the recipient at the car dealership. Otherwise, the advisor will need to assume how the automobile may look like in the car dealership, or extrapolate from that same brand and model seen in the street, along with hypotheses about how the salesman may attempt to persuade the recipient. In either case – the advisor being able to experience the decision context or not – the advisor will make assumptions about what elements form the decision context of the recipient. Members of the other set of assumptions together amount to, roughly speaking, a model of man, that is, the advisor's conception of the recipient. This conception equates to assumptions about what information the recipient may have and how he may be using that information when choosing. Returning to the automobile example, the advisor may assume, e.g., what vehicle qualities the recipient may pay particular attention to. Based on such assumptions, the advisor may then recommend a particular brand and/or model over another, or, instead of immediately offering an alternative course of action, the advisor may recommend other qualities that the buyer should take into account.

The rough formulation of the advisor's problem is this: after the advisor forms the two sets of assumptions, he needs to design such advice, which would lead the recipient to choose the course of action favored by the advisor. From the assumptions on how the recipient decides, the advisor designs his advice as some information that will, once communicated, change how the recipient perceives the decision problem. That change in perception will, hopefully, be such as to point the recipient in the direction intended by the advisor. A classification of advice can be made on the basis of what the designed advice will be targeting at the recipient. Take the car purchase example again: suppose that the advisor is convinced that the model B is the best choice for the recipient, but assumes that the recipient will choose a model A. If the advisor assumes that the recipient has never seen the model B, he will design different advice than if he assumes that the recipient knows of B, but prefers A because of some quality, e.g., the engine size. In the former case, the advisor will design advice that introduces the recipient to B, and only then establish, through communication, whether the recipient prefers B over A or the other way around. In the latter case, there is no need to introduce the recipient to B, but rather, design advice that will aim to influence the relative evaluation of A to B that the recipient seems to make. That advice is designed to suit a purpose has as a consequence that advice will be tailored to its target, and we see from the example that targets can vary. It consequently becomes interesting to classify advice according to its target.

If we can classify advice depending on its target, we need to determine the range of plausible targets. Now, remember that the targets are in the eye of the advisor, that is, the range of plausible targets in a particular advice-giving setting will depend specifically on the assumptions that the advisor makes about the information that the recipient has, and how the recipient uses that information. If we abstract from the particular information in a particular decision setting, we conclude that the advice can be designed to target different classes of information. The classes of information in turn are defined by what we called above the model of man that the advisor has

of the recipient. If the advisor assumes that the recipient decides by maximizing the expected utility, then the advisor is effectively assuming that the recipient is the rational economic man, and consequently decides according to the model of rational choice, which in turn hypothesizes very specific classes of information that are relevant to the decision problem. If we know the model of rational choice, we can classify kinds of advice depending on the information that advice can target in that model of choice. If the model of choice says that the decision-maker, here the recipient, estimates, for each alternative the probability for its effects to occur, and the utility of those effects, then we can have the following classification: advice that aims to change the probability estimate, advice that aims to change the utility estimate, advice that changes the set of alternatives, and advice that changes the set of effects of alternatives. We see that the range of targets will depend on the model of man that the advisor has of the recipient.

Taken in its very broad sense, the term *model of man* refers to an individual's concept of a human being. Different fields have constructed their own local conceptions, with which they highlight some properties of a human that they are particularly interested in. In biomechanics of humans, for instance, the interest is in understanding the mechanics of human movement, such as walking, jumping, and running (Alexander 2003). As there is no interest for, say, why the person is walking somewhere, but only in how walking happens from a mechanical standpoint, there is no need in a model of man in biomechanics to incorporate information about the presumed motives of a person. For a psychologist, the model of man will have a different focus, aiming perhaps to describe how people may learn from past experience, what may influence such learning, or how attitudes may affect behavior (Bandura 1974). For the economist and the manager, a model of man will state what an individual may take into account when choosing (Mueller 2004), such as alternatives, preferences, and outcomes, and what rule he may be applying to rank alternatives (e.g., maximize own profit or utility).

Knowledge of, say, models in human biomechanics can help in advising someone on how to walk or run. It is also evident, however, that we cannot give a general classification of advice if we take the model of man from biomechanics; if we do so, the classification will apply to advice given only with regards to the decision situations that fall within biomechanics (e.g., choosing how to run). When an individual is advised to walk differently than he currently does, knowledge in biomechanics proves relevant in choosing the content of advice; the target of advice – in terms of the information that the recipient has and the way he uses it – still remains open. While the recipient of advice is indeed making a choice with regards to biomechanics, he is facing a decision situation. As economics and management produce general models of the choosing individual, they remain uninterested with the specifics of the decision situation, i.e., whether it is buying an automobile, changing the way one runs, and so on. It thus seems plausible to look at main models of choice to see how advice can be classified, depending on the element of the model of choice (e.g., alternatives, preferences) it will target (e.g., advice on what alternative to scrap, or what to prefer). With this in mind, we will go over several models of choice often cited in economics and organization sciences. To be interested, as we are in the ingredients of these models, is to seek for each model its ontology.



## 5.2 Perfect and Bounded Rational Choice

The notion of *methodological individualism* refers to the claim that social phenomena should be explained in terms of how they result from the actions of individuals. If one is interested for instance in why a corporation went bust, he should look into the actions of the decision-makers, whose choices could have significantly affected the behavior of the employees and others, say, its suppliers, buyers, and individuals acting within regulatory bodies. Introduced by Max Weber, the nineteenth to twentieth century sociologist, methodological individualism refers to a method of inquiry that stands in contrast to other approaches to the explanation of social phenomena, such as sociobiology, psychoanalysis, or generalizations from purely statistical observation (Heath 2009). In case of sociobiology, roughly speaking, explanation of individual and collective behavior will revolve around the biological factors (e.g., genetics) and their relevance in the Darwinian evolutionary process; psychoanalysis would emphasise the role of personality traits and sexuality, while purely statistical observation would hypothesize only from measurable behavior.

Neoclassical economics relies on the method of methodological individualism. It consequently needs a model of man, which hypothesizes how an individual chooses. In having such a model, it makes the individual predictable when placed within some context of decision. The model is thus at the basis of all prediction that may be done via any model falling within neoclassical economics, either at the level of a single individual or that of a group.

According to the model of man in the mainstream of neoclassical economics, the individual's basic and general aim is to maximize his benefits and minimize his costs. To make the optimal choice, the decision-maker has perfect information, that is, knows *all* alternative courses of action, all consequences of all alternative courses of action, the likelihood that these effects will be brought about as a result of choosing any of the alternative courses of action. Although this already is much, it is not enough, for it is also necessary to compare alternative courses of action in terms of their desirability, to establish preferences between them. Hence, it is also said that the individual can – to use a computing metaphor – process all of that information.

It is not difficult to find an angle of attack on how perfectly rational choice conceives the decision-maker. It is not unsurprising then, as we already noted earlier that much criticism, adaptation, and revision has been produced mostly since the 1950s, when, among others, the very visible game theory was introduced by John von Neumann and Oskar Morgenstern (von Neumann and Morgenstern 1956). Despite the many variants of rational choice models already in mid-twentieth century, Herbert Simon noted that “[m]odels of rational behavior — both the global kinds usually constructed, and the more limited kinds ... — generally require some or all of the following elements” (Simon 1955, p. 102), which together form the ontology of the perfectly rational model of choice:

1. *Behavioral alternative.* A set of behavioral alternatives, say  $A$ , is available to the decision-maker. The decision-maker “considers” (i.e., chooses from) a subset  $\overset{\circ}{A}$  of all possible alternatives, i.e.,  $\overset{\circ}{A} \subset A$ .

2. *Decision outcome.* A decision outcome designates states of affairs that are brought about by following corresponding behavioral alternatives. There is a set, say  $S$ , of decision outcomes.
3. *Payoff function.* A payoff function, say  $\mathcal{V}$ , relates individual outcomes  $s \in S$  to individual “values” or “utilities”  $u \in U$  that the decision-maker attributes to that outcome. A payoff function can be replaced by an order on members of  $S$  to indicate the order of preference instead of associating utility values to outcomes.
4. *Probabilistic outcome function.* An outcome function relates individual behavioral alternatives  $a \in A$  ( $a \in \mathring{A}$ ) to individual outcomes  $s \in S$ . When there are more than one possible outcome for a single behavioral alternative,  $S_a$  designates the set of possible outcomes for the alternative  $a$ . A probabilistic outcome function relates  $a \in A$  ( $a \in \mathring{A}$ ) to individual pairs  $(s, P_a(s))$ , where  $s \in S$  and  $P_a(s)$  is the probability that  $s$  ensues if  $a$  is chosen;  $P_a(s)$  is such that  $\sum_{S_a} P_a(s) = 1$ .

Given the elements above, the problem that the perfectly rational decision-maker faces is this:

**The Perfectly Rational Choice Problem.** Given known behavioral alternatives  $A$ , decision outcomes  $S$ , payoff function  $\mathcal{V}$ , probability distribution functions  $P_a(s)$  for all  $a \in A$  and  $s \in S$ , find the alternative that maximizes the payoff.

Various ways to resolve this problem, that is to select the appropriate course of action, can be defined, each amounting to what Simon calls “classical” rationality:

- *Max–min rule.* If it is assumed that whatever alternative is chosen, the worst possible outcome will ensue (i.e., the smallest  $\mathcal{V}(s)$  for  $s \in S_a$  will be realized), then select  $a$  with as large as possible worst payoff: doing the chosen  $a$  should give the outcome with the payoff  $\text{Max}_{a \in A} \text{Min}_{s \in S_a} \mathcal{V}(s)$ .
- *Probabilistic rule.* Choose  $a$  that maximizes the expected value of  $\mathcal{V}(s)$  for the probability distribution  $P_a(s)$ : doing the chosen  $a$  should give the outcome with the payoff  $\text{Max}_{a \in A} \sum_{s \in S_a} \mathcal{V}(s) P_a(s)$ .
- *Certainty rule.* If it is certain that each  $a \in A$  maps to a specific  $s_a \in S$ , select  $a$  with the largest payoff: doing the chosen  $a$  should give the outcome with the payoff  $\text{Max}_{a \in A} \mathcal{V}(S_a)$ .

Simon’s behavioral model of rational choice challenges such an understanding of decision behavior, by arguing that the inherent limitations of the decision-maker restrict the information that can be considered and processed when making a choice: “actual human rationality-striving can at best be an extremely crude and simplified approximation of the kind of global rationality that is implied, for example, by game-theoretical models” (Simon 1955, p. 101). Simon’s model introduces simplifying assumptions or, equivalently, approximating procedures to global models that describe perfectly rational decision-making. To introduce simplifications in perfect rationality, to render it bounded, Simon does not change the ontology given above. Rather, what changes are the problem formulation and the resolution procedures. He deemed it more realistic to assume that the decision-maker uses simple payoff functions – e.g., all outcomes may be split into satisfactory or unsatisfactory

ones, instead of associating a numeric value to each possible outcome. The second simplification arises when mappings from behavioral alternatives to outcomes are unknown and finding them is costly. Consequently, information-gathering steps need to be introduced in the decision-making procedure, and it becomes important to determine how much resources to invest in such steps. The third simplification is to rely on a partial ordering of payoffs, instead of assuming that the individual totally orders all considered payoffs. These simplifications lead Simon to introduce the following resolution procedure:

**Satisficing.** If there is a threshold payoff  $u_{min}$ , i.e., a minimal acceptable payoff, then: (1) search for a set of possible outcomes  $S' \subset S$  such that  $\mathcal{V}(s)_{s \in S'} \geq u_{min}$ ; and (2) search for a behavior alternative  $a \in \overset{\circ}{A}$  whose possible outcomes are all in  $S'$ .

Satisficing seeks not the *optimal* behavioral alternative, but one that is merely *good enough with regards to the minimal acceptable payoff*; in doing so, a complete ordering of payoffs is not necessary, and the procedure can be applied without assuming that either the individual somehow has all information, or that he can spend any time he wishes to collect information. The satisficing rule resolves a weaker, though undoubtedly more realistic decision-making problem than that of the perfectly rational choice. The weaker problem can be formulated as follows:

**The Bounded Rational Choice Problem.** Given an approximate payoff function  $\mathcal{V}$ , and a minimal acceptable payoff  $u_{min}$ , find possible outcomes  $S' \subset S$  such that  $\mathcal{V}(s)_{s \in S'} \geq u_{min}$ , and find a behavioral alternative  $a \in \overset{\circ}{A}$  whose possible outcomes are all in  $S'$ .

Within the limits of the resources available for decision-making, the individual who satisfices will set a threshold payoff (i.e., the minimally acceptable payoff), then look for the first alternative with that results in an outcome valued above the threshold. It should be clear that this is very different from optimization: if the individual wishes to find the optimal, or the best of all alternatives, he must first find all alternatives, while he should of course start by assuming that he actually can find all alternatives.

What do we learn for a classification of advice from these expositions of rational choice theories? To advise an individual who is assumed to command either perfect or bounded rationality is to produce advice of one of six kinds. To the extent that the ontology of rational choice bears six concepts – namely: (behavioral) alternative, (decision) outcome, payoff function, probabilistic outcome function, perfectly rational choice problem, and resolution procedure – the decision-maker can be advised on any of the six. We would consequently specialize the concept of advice on six other concepts, e.g., advice on (behavioral) alternative, advice on (decision) outcome, and so on. Advice may thus recommend a reevaluation of a utility value for the outcome of an alternative, or recommend the use of a different resolution procedure (e.g., Max–min instead of the Certainty rule). When advice concerns the problem itself, it may aim to clarify to the individual that he is indeed facing this and not another kind of problem; somewhat comically, a person puzzled by some complex decision problem could thus be advised that he is in fact very much in command of all the information he may need to choose, since he is a perfectly rational decision-maker.

There is very little interest in advising the individual who meets the desiderata for perfect rationality. While we can identify kinds of advice from the ontology of rational decision-making, why bother instantiating the concepts of that ontology when the recipient already knows all that he needs to know? The paradox is apparent: if any advice given to the perfectly rational recipient is not already known to the recipient, then the advisor is somehow “more perfectly” rational than the recipient. To argue that the advisor may facilitate the acquisition of information about, e.g., the alternatives or outcomes, also misses the point: perfect information *is* assumed at the outset. Perfect rationality indeed makes communication irrelevant and consequently all advice irrelevant to choice. It is thus not unexpected to see that John Nash assumed no communication between competing players in a game theoretic setting (Nash 1951, p. 286): each player chooses by anticipating the actions of the other players, but does not communicate with them.

As soon as we admit that the decision-maker is somehow imperfect, advice giving becomes of interest. If rationality is bounded in Simon’s way, the decision-maker will be spending some time to structure his decision problem, i.e., identify some (but not all) alternatives, identify some outcomes thereof, estimate probabilities, evaluate outcomes, and so on. Advice may be given during the time between the moment the decision-maker becomes aware that he should choose, and the choice itself. Since he cannot know fully, as under perfect rationality, all potentially relevant information to his decision problem, the advisor may make suggestions that would perhaps add to or change the existing set of information – on alternatives, outcomes, probabilities, payoffs, etc. – that the decision-maker has acquired himself. There is a large body of work in economics on what/how information can be passed to decision-makers within this window.

A broad distinction can be made between research into cases where an action other than usual conversation conveys advice to the decision-maker, while other research has been interested in advice giving via informal and usual conversation. Michael Spence’s notion of *signaling* refers to observable actions that a seller in a transaction may undertake to convince the buyer in the value of the products or services being sold. The classical example is the signaling that job applicants perform to the attention of employers in a job market (Spence 1973). Since the employer is, at the time of hire, uncertain about the value he can obtain from the candidate, the wage can be assumed to depend on the perceptions he has of the candidate. The perceptions are in turn determined by the observable qualities of the candidate, such as education and previous work. Since the perception of these qualities enters the decision-making of the employer, the candidate designs advice by modifying his own qualities. Being a rational decision-maker (i.e., willing to maximize own benefits and minimize costs), the candidate is assumed to “choose” a level of education that gives him the biggest difference between the cost of education and the expected wage. The main interest in signaling literature that followed Spence’s contributions remains the relation between signaling on some market of interest and the efficiency of that market.

The impact that informal conversation between players of a game can have on the set of outcomes of the game (i.e., combinations of players’ individual decisions) is

studied in the literature on *cheap talk*: “Cheap talk is just that: cheap — neither costly nor binding; and talk — not some roundabout form of communication, like mediation. Unlike ‘signaling’, cheap talk — plain conversation — is ‘payoff-irrelevant’ [...] [The decision-makers in the game] may or may not tell the truth, and may or may not believe each other” (Aumann and Hart 2003, p. 1619). There are no general results for how cheap talk will affect the outcome of a game: it need not lead to an efficient outcome, although it may do so. The more general point is very much in line with the ideas on advice here: cheap talk (some of which may be advice) can change the players’ understanding of the game, thereby influencing the outcomes of the game. We mentioned in Chap. 2 Schotter’s experiments on the use of advice in intergenerational ultimatum games, where advice did affect the thinking and choices of the players who received advice, and ultimately, the outcome of the game. It should be clear that the game itself in that case was defined and advice could not change it (e.g., change the number of alternatives, or players). There is no reason of course to believe that a situation resembling a game in actual settings should always have a fixed structure.

### 5.2.1 *Ontology of Decision Analysis*

Decision analysis in management science comes mostly from Ronald Howard’s contributions in the 1960s (Howard 1966) and later developments by Ralph Keeney and Howard Raiffa (Raiffa 1968; Keeney and Raiffa 1976). Decision analysis starts from the axioms of perfectly rational choice and aims to be prescriptive. It takes the form of a methodology to fulfill its normative aim. At the first step, the decision problem is structured, that is, the following are identified:

1. *Alternative*. At least two alternatives — i.e., alternative courses of action that resolve the problem of interest — need to be identified. Assume that there are  $J$  alternatives:  $A_j, j = 1, \dots, J$ .
2. *Objective*. One or more objectives — i.e., goals to achieve, purposes to fulfill — are identified, and each is refined into more specific and measurable objectives, so that satisfying the latter equates satisfying the former. Assume that there are  $n$  lowest-level objectives:  $O_i, i = 1, \dots, n$ .
3. *Attribute*. An attribute measures the degree to which an objective is achieved. Attributes are identified and associated with all the lowest-level objectives. Assume that there are  $m$  attributes:  $X_k, k = 1, \dots, m$ ; let  $x_k$  be a specific level of  $X_k$ .

The second step assesses the possible effects of different alternatives. The aim is to identify, for each possible alternative, the probability of each possible consequence, so that the following information is sought.

4. *Probabilistic consequence function*. This function is a probability distribution function mapping values of attributes to the probability of their occurrence. It is

defined for each alternative. Let  $p_j$  denote the probability consequence function for the alternative  $A_j$ .

In the third step, the desirability of the alternatives is evaluated, that is, utility functions are elicited:

5. *Utility (equivalently: value) function.* This function maps attribute values of an alternative to a value that is interpreted as providing the overall desirability of the alternative. The general form of the utility function is:  $u(x_1, \dots, x_m) = f(u_1(x_1), \dots, u_m(x_m), c_1, \dots, c_m)$ , where each  $u$  is a single-attribute utility function and each  $c$  is a scaling constant. The scaling constants indicate the relative desirability of an attribute compared to other attributes.

The fourth step of decision analysis uses the collected information to rank alternatives and identify the most appropriate one. Before the decision rule of the fourth step is reiterated herein, it is relevant to place the previous steps in perspective by summarizing the very problem of decision analysis. Applying decision analysis amounts to the instantiation of the problem defined as follows.

**The Decision Analysis Problem.** Within the resources allocated to the achievement of the objectives, refine the objectives, identify the alternative courses of action that are assumed to satisfy the lowest-level objectives, and define the attributes that quantify the level to which the objectives are achieved. Elicit then the probabilistic consequence functions and the utility functions. Finally, find among the considered alternatives that which gives the maximal expected utility.

The problem of decision analysis is to rank alternatives by synthesizing their relative cumulative utility; that cumulative utility is the decision criterion is due to the axioms of rational choice. The highest ranking alternative should be chosen as the most appropriate one within the set of considered alternatives. The problem of decision analysis appears mid-way between that of perfectly rational choice and the Simon's bounded rational choice: while perfectly rational choice assumes available information on alternatives and limitless resources for the comparison of alternatives, decision analysis assumes that the variety of identified alternatives – and the detail thereof – will ultimately depend, among others, on the resources invested in the application of the decision analysis methodology, the creativity of the participants, and the inherent features of the problem. It is, however, also apparent that a toy decision problem in decision analysis – in which an exhaustive and complete set of possible alternatives and their likelihoods can be determined and in which relationships between alternatives and outcomes can be precisely known within some realistic resource bounds – can be readily formulated as the problem of perfectly rational choice. As far as Simon's satisficing is concerned, the methodology of decision analysis is flexible enough to allow the application of the satisficing rule. Returning to the fourth step of decision analysis, it amounts to the application of the following resolution rule.

**Maximize Expected Utility.** Assume a set of identified alternatives, objectives, attributes, probabilistic consequence functions, and utility functions. Following the axioms of perfectly rational choice, compute for each alternative its expected utility, that is,  $E_j(u) = \int p_j(x)u(x)dx$ , and choose the alternative yielding the highest expected utility.

The ontology of decision analysis only slightly differs from that of perfectly rational decision-making. The differences are due to the insight gained through the research on normative procedural aspects of decision-making. Namely, objectives and attributes thereof are recognized explicitly, thus giving a more elaborate account of the decision setting than when outcomes are considered. Also, it is explicit that attributes may be of different importance to the decision-maker. Note that these divergences augment the ontology of rational choice. It is not difficult to observe that outcomes equate with tuples of values of attributes, while objectives add an additional layer that is of methodological interest. Also note that the maximization of the expected utility is a resolution procedure that applies to the problem of perfectly rational choice, with the apparent caveat that not all possible alternatives, outcomes, and other information are known. Those that are known are those identified by applying the steps of decision analysis.

Just as the ontology of perfect and bounded rational choice gave a candidate classification to specialize the concept advice, so does decision analysis provide such a candidate. To advise the individual who obeys the tenets of decision analysis is to produce advice on his view of any of the alternatives, objectives, attributes, probabilities of the attributes to occur, the form of the utility functions, or the choice of scaling constants. Beyond the elements of the decision problem, advice can be offered on the methodology to follow toward the structuring of the decision problem itself; decision analysis is, of course, itself and in its entirety advice on how to conceptualize a decision situation (Howard 1980), that is, an elaborate recommendation on how to proceed to distinguish noise from what decision analysis assumes are relevant elements of the decision problem. If there ever was a general body of knowledge for professional consultants, it was that of decision analysis.

### 5.2.2 *Ontology of Choice in Organized Anarchies*

As long as the sole victim of an individual's choices is himself, how he chooses is his own business. There is, however, in most choices some collateral damage. This is especially the case in organizations, where choices not only do affect others, but are made through interaction with them. To understand how advising an individual within an organization may differ from advising a perfect or bounded rational chooser, we need to look into models of organizational decision-making.

As Ann Langley and her colleagues observe "[a] large proportion of [...] research on organizational decision-making followed Simon's lead. Underlying much of this work has been the view of decision-making as a boundedly rational process converging sequentially from the stage of problem definition towards that of final choice." (Langley et al. 1995, p. 262) Focus has consequently been placed on the procedural aspects of boundedly rational choice, while remaining within the boundaries of the ontology initially defined for perfectly rational decision-making, which remains in Herbert Simon's ontology for bounded rational choice, and is the foundation of decision analysis.



If we seek other perspectives on choice in organizations, then a candidate is Michael Cohen, James March, and Johan Olsen's "garbage can" model of organizational decision-making (Cohen et al. 1972). There, the focus is on choice when preferences vary and may be inconsistent, methods for the resolution of the organization's problems are not standardized, and various actors can take part in decision-making. As a descriptive model at odds with the idea that joint decisions happen along the lines of decision analysis, the garbage can assume less of a structured and systematic behavior when choices are made. The interplay among participants who promote personal interests, e.g., through bargaining or coalition building is emphasized. While decision-making does appear chaotic, it remains "channeled by structural features of hierarchy and functional specialization" (Pinfield 1986). In other words, authority plays a role, as does more generally an individual's formal position within the organization. Cohen and colleagues see a decision as an outcome of several relatively independent streams within the organization. A stream involves a sequence of the instances of one of the following concepts:

1. *Problem* designates a situation that is undesirable to the members of the organization or to its stakeholders. Resolving a problem requires the investment of a certain degree of energy from the participants.
2. *Solution* is a means needed to resolve a problem. "Despite the dictum that you cannot find the answer until you have formulated the question well, you often do not know what the question is in organizational problem solving until you know the answer." (Cohen et al. 1972, p. 3) Solutions are thus available in an organization regardless of whether the organization has the corresponding problems: a solution need not be found after the problem is found, instead, a problem may be identified based on the availability of the solution.
3. *Participant* is an actor within or from the outside of the organization that can invest energy in the resolution of problems. The energy they may invest varies over time.
4. *Choice opportunity* is an occasion "when an organization is expected to produce behavior that can be called a decision." (Cohen et al. 1972, p. 3). In other words, a choice opportunity is a time when a solution is matched to a problem.

Choice opportunities occur from the mixture of problems, solutions, and participants. Such mixtures will less frequently produce solutions to problems; instead, it is more common that problems get attached to participants. While problems will surface at various choice opportunities, they will rarely be resolved by a choice of a particular solution: problems can be dealt with without explicit choice, and choices can be made without addressing explicit problems. In this sense, Cohen and colleagues see the notion of decision as a post-factum construct identified by participants in decision-making or those studying the organization (Pinfield 1986). The decision problem of such "organized anarchies," that is, organizations to which the garbage can model of choice applies, can be defined as follows.

**The Problem of Choice in Organized Anarchies.** Given unpredictable streams of problems, solutions, participants, choice opportunities, participants' availability constraints,



and constraints on participants' eligibility to take part in decision-making, the eligible participants need to find solutions that fit to problems associated to each given choice opportunity.

The decision procedure that responds to the problem is, from the perspective of rational choice, just as chaotic as the problem:

**Garbage Can Resolution Procedure for the Problem of Choice in Organized Anarchies.** Associate eligible and available participants to the choice opportunity, so that they can invest effort in finding solutions to the problems associated with the choice opportunity. Until the problem becomes irrelevant and while it is still unresolved, wait for the problem to become associated to a new choice opportunity, then retry to resolve it.

Decision-making in organized anarchies still articulates similar notions to those of perfectly rational decision-making: regardless of how things happen in the organized anarchy, there is still at least one possible choice at each choice opportunity (which may simply amount to delay the decision), that is, one possible behavioral alternative, choices still have some payoff for each participant (even though they differ among participants), actions still have expected outcomes, even though relationships between the former and the latter may be unintelligible within the limits of availability and eligibility of participants. Each participant behaves as a boundedly rational decision-maker, i.e., satisfices when choosing. However, the combined behavior of different decision-makers does not resemble a structured decision process; in other words, together, they end up neither with satisficing nor optimizing *joint* behavior. In the extreme case, they resemble "collections of choices looking for problems, issues and feelings looking for decision situations in which they might be aired, solutions looking for issues to which they might be an answer, and decision-makers looking for work" (Cohen et al. 1972, p. 1).

Two classifications of advice are interesting when recommending courses of action to members of organized anarchies. In so far as each of these members chooses along the model of bounded rational choice, he can be advised on alternatives, payoffs, probabilities, preferences, and so on; advice can be specialized thus along the ontology underlying Simon's model of the decision-maker. Another classification could distinguish kinds of advice depending on what element they target in the problem of choice in organized anarchies. We could thus have advice on problems, solutions, participants, and choice opportunities. The two classifications would not be incompatible, though they seem to overlap in some concepts: e.g., behavioral alternatives appear similar to solutions.

### ***5.2.3 Ontology of Intervowen Organisational Choice***

Conceptualizations of individual or joint choice outlined above – perfect and bounded rational choice, decision analysis, garbage can model – are, for better and worse, the foundations of research on individual and organizational decision-making and remain dominant views on how choices are and/or should be made.

Ann Langley and her colleagues' conceptualization discussed at present has had less time to spread and be subjected to critique. Nevertheless, it provides strong and plausible arguments for the need to take another view on decision-making, that is, advance toward a richer understanding of how decisions are made in organizations. They voice three concerns: (1) the concept of "decision" may most often be no more than a construct in the eye of the observer and attempting to do pin them down when studying organizational decision-making "may distort our perceptions of how action really occurs in organizations" (Langley et al. 1995, p. 270); (2) "the view that decisions unfold in a sequential pattern, oblivious of individual differences and divorced of human emotion and imagination" (Langley et al. 1995, p. 264) is overly simplistic – in particular, the role of affect should be given a more prominent position in conceptualizations of decision-making; and (3) "the assumption that decision processes can in fact be isolated from one another and from much of the collective reality that is organization" (Langley et al. 1995, p. 264) does not hold – instead, interrelations and linkages between choice processes need to be accounted for to understand organizational decision-making. Motivated by these observations, Langley and colleagues draw on a wide and deep body of theoretical and empirical research to advance their own conceptualization, summarized in the following:

decision-making comes to be seen here as a complex network of issues involving a whole host of linkages, more or less tightly coupled. Periodically decisions emerge from this network, or at least actions, driven by insights as well as various affective factors in addition to the cerebral rationalities of the actors. The apt analogy here is the moving stream, the context in which the issues float along, sometimes getting washed up on shore as actions, sometimes sinking and disappearing, and often bumping into each other with the effect of changing another's direction, slowing one down, speeding one up, joining two together, or having a single issue burst into several new ones. (Langley et al. 1995, p. 275)

The conceptualization that can be deduced from their discussion involves the interplay of the following concepts and relationships:

1. *Issue*. An issue designates something that is undesirable from the perspective of the participants in the organization or its stakeholders, and thereby would eventually require action to be taken to bring about change that will make the issue no longer relevant.
2. *Decision*. A decision is an explicit choice of a particular course of action in relation to one or more issues.
3. *Issue stream*. An issue stream is a representation of an issue over time. When actions are performed in relation to an issue, these actions are associated with particular points on an issue stream.
4. *Decision linkage*. Given that a decision is an explicit choice of action with regards to an issue, a decision is associated with a particular point on the issue stream. Decisions on one or more issues affect future decisions on the same or other issues. A decision linkage captures this influence in the form of a linkage between decisions on the same or across issue streams. There are three kinds of decision linkages:
  - (a) *Sequential linkages*. A sequential linkage exists between two decisions taken at different times on the same issue.

- (b) *Lateral linkages*. A lateral linkage exists between decisions on distinct issue streams, whereby the decisions share the same resources (e.g., budget, time) or the same organizational context (e.g., same people, culture).
  - (c) *Precursive linkages*. A precursive linkage exists between decisions if the decision on one issue can affect the premises on subsequent decisions on other issues (e.g., a decision enables other decisions).
5. *Issue network*. Issue streams, decisions, and decision linkages together form an issue network.

While their preoccupation is essentially the same as those of Simon, Cohen and colleagues, and decision analysts, the approach is different. The conceptualization makes no explicit assumptions about the problem and decision rule to apply. It is a more abstract description of the organizational decision-making process, in which more or less rational decision-making can take place. It is thus not unreasonable to assume that within the setting that Langley and colleagues define, various choice behaviors can be observed, each more or less similar to that of rational or boundedly rational choice, while affect, context, and the interdependence of issues is assumed to mediate all choice behavior. It is that mediation that makes the rational and boundedly rational conceptualizations overly simplistic when asked to explain decisions within issue networks.

Issue networks cannot exist without continual and repeated social interactions via the communication among participants, whereby the content of the communication conveys information that invariably acts as an input to decision-making; invariably indeed, because the individual participant is hardly isolated from other participants, from what they communicate, do, and from the effects they produce by doing. Information that is exchanged enters into the deliberation of the participant about issues and alternative courses of action, and thereby affects his choices made across issue streams. Organisational decision-making can thus hardly be seen as a solitary activity. It is a social one, in which all participants have a role to play in addressing organizational issues. Now, while in a machine-like bureaucracy (Mintzberg 1979) a course of action would not be taken – to use Douglas Adams' caricature from his *Hitchhikers Guide to the Galaxy* – “without orders signed in triplicate, sent in, sent back, lost, found, queried, subjected to public inquiry, lost again” (Adams 1979) and hence, formally recorded and acknowledged, the roles and effects of decision-making are not necessarily anchored only in the formal structure and authority, or should have a tangible representation. Communication of information that enters decision-making is thus not unidirectional, from the participants to a formal decision-maker. In other words, many decisions are continually being made, by various participants, and each may influence that of another. Just as the manager would affect the managed, so does the latter influence the former, both via communication. Every participant decides on a variety of issues, whereas some do have more influence in shaping issue streams by their formal authority or visibility in the organization. Organisational decision-making is thus decentralized, and the path that the organization takes is shaped by the mixture of its various participants' choices and actions in response to the issues: each participant's choices of action,

effects of actions, and the information conveyed by communicating with other participants enters into account in the decision-making of those others, with whom the individual interacts. Not much should be expected in terms of coherence or precision of information that participants communicate. They may hold false or contradictory information, they may agree with some general principle but disagree with some specific aspect thereof, they may come to agree with it later on, or reject the general principle altogether (Kuwada 1998). Along this same line of reasoning is that the informal or formal status of an expert within an organization gives no guarantee that the information provided by that participant is relevant.

In summary, the organization is thus a social formation within which interdependent participants continually interact. They are interdependent simply because each one alone cannot resolve issues, thereby requiring some form of interaction with others, or that the consequences of the actions of some affect others (i.e., an issue concerns more than one actor). Their actions and interactions convey information. This information may be intentionally or unintentionally directed toward other organizational actors, taking the form of advice. Advice enters in the decision-making process of each individual actor to whom the information has somehow been brought to attention. The organization thus effectively amounts to many interdependent decision-makers distributed over common issues. Most importantly, the decision rationale of each actor is shaped by the information acquired via communication, and by the information that the actor generates by analyzing the acquired information. All of this information forms the context of that actor's deliberation before choice, whereby the rationale structures this information.

What kinds of advice would be dispensed within organizations resembling those akin to issue networks? It should be a straightforward matter by now to see that, as before for rational choice, advice can be specialized along the concepts forming the ontology of the model of choice. For issue networks, we could distinguish advice on issues, from advice on how diverse decisions relate and impact other issue streams. Since the model is one of organization, not of man, various other kinds of advice could remain useful. The decision analyst who would accept Langley and colleagues' model of organization can still remain with a classification of advice that follows the ontology of decision analysis. What issue networks add is the emphasis on the evidence that many choices are being continually made within an organization, which in turn requires any choice to consider not only the effects within its issue stream, but also on those outside of it.

### **5.3 Intolerance for Substitutes**

If advice is specialized according to its target, as we illustrated in the past section, then the ensuing ontology of advice is determined by the ontology of the model of the decision-maker. To prefer then, as many have the model of the perfect or bounded rational economic man, is to distinguish advice on alternatives, from advice on payoffs, from that on the probabilistic outcome function (i.e., advice

on probabilities), and so on. Such an ontology of advice mirrors the ontology of information that the decision-maker is assumed to account for when choosing.

There is a very simple, though not obvious, reason why *any* model of choice that integrates probabilities and utilities will yield a dysfunctional ontology of advice. It is this: because a model of choice relies on probabilities and utilities, it separates the rationale behind the probability and utility estimates from the formulation of the decision problem. Both a probability and a utility estimate do not and should not come out of thin air: there are reasons why a decision-maker would give one value  $x$  in the interval  $[0, 1]$  of real numbers as a probability of some outcome, instead of some other value  $y \neq x$  in that same interval. As such, probability is merely a crude and summary substitute, a formulation of the result of the thinking that the decision-maker had invested in assessing the information he has about the decision problem. A probability value is a quantity that – by convention in the rational models of choice – refers to a degree of uncertainty about a future occurrence of some event. Same observation applies, *mutatis mutandis* to utility: a utility value is a crude and summary formulation of the result of the decision-maker's reflections on the desirability of an outcome relative to other outcomes he has taken into consideration.

What does the rationale for probability and utility estimates have to do with advice? Suppose that an individual chooses by following the axioms and model of decision analysis, so that he has come up with some alternatives, probabilities, utilities, and so on; he has, in other words, instantiated the ontology of decision analysis to describe his concrete decision problem. If an advisor wished to recommend to this decision-maker a different estimate of the probability of some outcome, would the advisor simply suggest another probability value? If our ontology of advice followed the ontology of decision analysis, there would effectively be a kind of advice that targets probabilities estimated by the decision-maker, so that to advise on probabilities is indeed to recommend probability values. If the advisor does recommend a different probability value, why would the decision-maker accept it? The only case he should accept it is if the mechanism of consent obliges here to do so, and he does not wish to venture against it. In all other cases, if the decision-maker did invest some thought in his probability estimate, and if the advisor suggests another one, what ensues is a discussion of why the recommended estimate would be more appropriate than the initial one. The decision-maker may accept the recommended probability value, e.g., if the advisor shows a fault in the thinking and information that led the decision-maker to the initial estimate. But such a discussion will not involve advice about numbers, but advice about the information relevant to the decision-maker and its analysis. If advice on a probability estimate is simply another probability value, it will hardly be of any interest to the decision-maker, for he will presumably ask *why* that recommended estimate is any better than his own. As soon as the *why* is evoked, advice that will be given to the decision-maker will no longer be about the probability value itself, but about the rationale that led to that probability value and not another one. The point is that an ontology of advice should acknowledge this: namely, that advice that only speaks of a probability value (and not of why that value is a relevant estimate of uncertainty) is an ontology that leaves outside its scope the evidence that an advisor will aim to influence the probability

value itself, but the rationale of the decision-maker. Such influence does not happen by recommending *only* that a probability value  $x$  is more appropriate than a different probability value  $y$ ; rather, it requires that the choice of one value over another be explained to the decision-maker. Same reasoning and remarks apply for utility values.

The absence of concepts – in ontologies of decision-making based on expected utility theory – that would encompass the information and reasoning behind a decision-maker's probability and utility values makes these same ontologies poor candidates to guide the design of an ontology of advice. If they are used, the result will be an incomplete ontology of advice that misses entirely any advice that targets the information used in, and the process by which the decision-maker arrives at the probability and utility estimates.

The search for determinants of choice, variously called uncertainty, risk, desirability, context, and so on, is an ongoing one, and claiming a definite answer – with the now available knowledge in economics, psychology, and so on – would only signal a naïve understanding of human decision behavior. It should, nevertheless, also be clear that probability or utility values are very poor means to explain the rationale of some observed choice. If one does favor description or prescription via probabilities and utilities, then it is the rationale behind probability and utility values, along with that behind the definition of alternatives and outcomes that is a much more interesting explanation for the rationale behind a decision. This is not difficult to see: suppose that an alternative  $A$  is chosen as a solution to an instance of the problem of perfectly rational choice. The question of why choose the alternative  $A$  is answered by the resolution procedure applied to the decision problem: e.g., if the rule is the probabilistic one, the rationale for choosing  $A$  is that  $A$  maximizes the expected utility. One cannot be content with such an explanation. The question that ensues naturally is why does  $A$  rank highest; the immediate answer is that it bears the most favorable combination of utility and probability compared to the other alternatives. This again is uninteresting, since it states what is already apparent in the probability and utility assigned to  $A$ ; the truly interesting information are the arguments that led the decision-maker to associate the given values with the alternative  $A$ , and in case the choice was collective, or organizational, why other participants accepted those values. That information remains outside of the ontology of perfect and bounded rational choice, just as it remains outside of the ontology of decision analysis, garbage can model of choice, and the model of interwoven organizational choice. As this remains outside the cited conceptualizations, the decision-making information that they carry are thus separate from the information about the rationale from which the decision information originated in the first place. To know, e.g., that the choice of falsifying the financial reports had the highest expected utility for Enron executives is not enlightening, and neither is rationalizing via utility maximization the choice to disbelieve scientific research on climate change. This is not to say that utility maximization is not at work – whether it is, is actually irrelevant at present. What is affirmed herein, however, is that expected utility theory does not account for reasons for utility and probability estimates that a decision-maker provides, although it is these choices that ultimately determine the

optimal alternative. In a sense, the individual has chosen before he has even applied the decision rules in decision analysis, or any expected utility model: because a decision rule (e.g., max–min) simply aggregates the utilities and probabilities he has chosen, it is finally the rationale he has applied to arrive at these estimates that truly determines his choice. To be any relevant, advice must consequently target that rationale of the decision-maker. What model of that rationale we can offer, and thus suggest how to classify advice, is discussed in a subsequent section.

If advice cannot target a probability or a utility value alone, then we should ask what role remains for probability and utility in an ontology of advice? The rest of this section offers an answer.

### 5.3.1 *Probability Intolerance*

Ever since the early contributions in decision analysis, it was very clear to its proponents that it cannot be a proper description of how individuals do choose. This may seem paradoxical, since it starts from the traditional conception of the economic model of man, that is, the axioms of perfectly rational choice, and it would be expected that economics starts from something that even remotely fits *prima facie* evidence. To see this as a paradox is an error, since it is quite apparent from empirical evidence in even the simplest experimental settings that people do not obey these axioms (Slovic et al. 1977; Kahneman and Tversky 1979; Tversky and Kahneman 1981). It is not decision analysis that is descriptively misleading, but the axioms of perfectly rational choice when used for the analysis of decisions.

Aware that the axioms of perfectly rational choice are deficient, decision analysis does not say that these axioms *describe* individuals when they act as decision-makers, but that individuals *should* behave so when choosing. The aim is thereby no longer descriptive, but normative: “Decision analysis is a process” Ronald Howard argues “that enhances effective decision-making by providing for both logical, systematic analysis and imaginative creativity.” (Howard 1980, p. 4) The value thus seems to come out of the idea that the individuals involved in the decision situation need to exchange information in a step-by-step process, by instantiating concepts of the ontology of decision analysis. They are thus to be taught decision analysis and follow its suggestions when deciding. However limited the storage and processing abilities of individuals may be, they must aim to be economically rational within those bounds, while decision analysis can only make them “more” rational. In being rational, they will proceed step by step, from alternatives to a choice, juggling probabilities, and utilities along the way. So the story goes.

### Unexpected Conclusions

It is truly difficult to say what is more discouraging: that axioms that form the basis of decision analysis are descriptively misleading, or that decision analysis



sees this, but keeps insisting that people should still strive to choose as perfectly rational decision-makers. Ronald Howard cites Paul Slovic, Baruch Fischhoff and Sarah Lichtenstein's survey of the challenges to expected utility theory, or here, the axioms of perfectly rational choice:

The major advance in descriptive research over the last five years [i.e., first half of the 1970s] has been the discovery that people systematically violate the principles of rational decision-making when judging probabilities, making predictions, or otherwise attempting to cope with probabilistic tasks. Biases in judgments of uncertain events are often large and difficult to eliminate. The source of these biases can be traced to various heuristics or mental strategies that people use to process information [...] In the final discussion, a strong case is made that judgmental biases affect important dimensions in the real world; numerous examples are provided. (Slovic et al. 1977 cited in Howard 1980, p. 9)

As many others, Howard's conclusion from this is not that there is something wrong with the axioms of decision analysis, but quite the opposite:

As the degree of uncertainty goes up experimental subjects [i.e., decision-makers whose choices have been observed in controlled laboratory settings] begin to form false hypotheses and to retain them in the face of contrary evidence. [...] One could easily believe that human beings have very little inherent ability to handle uncertainty. People seem to have no intuitive idea of how to update their beliefs in the face of new evidence or of how the size of an experiment affects the inference that may be drawn from it. [...] I believe that in dealing with uncertainty the human being needs an instrument — probability theory, and that he will never be able to perform well in an uncertain environment without this instrument. (Howard 1980, pp. 12–13)

His conclusion can be rephrased as this: because people systematically misjudge (i.e., badly estimate) probabilities (as Kahneman and Tversky showed), they erroneously handle uncertainty, and consequently choose badly. It is not that this bad handling of probabilities leads to bad outcomes, but that the very method that people intuitively apply when deciding is dysfunctional.

Now, let us take a step back here, and see his argument for what it is: *if* one takes probability theory as a proper instrument to handle uncertainty, *then* people who cannot handle probability (those uneducated in decision analysis, among others) cannot handle uncertainty. There is a nuance here that must not go unnoticed, namely, that his conclusion is a conditional, where the condition is the acceptance of probability as a proper tool to handle uncertainty. If we can conceive of another way to handle uncertainty, then perhaps people are/should be using that one, and not probability theory. Perhaps the decision analyst sees errors only because he looks at people's choices thought this instrument, probability. Would the decision analyst still see errors if he was using another instrument?

It is *prima facie* evidence that most people are capable of deciding for themselves and make choices that are good enough to take them through life without too much harm. Life expectancy could not otherwise have moved from the *world average* of about 45 years of age in the mid-twentieth century, to 65 years at the outset of the twenty-first century, and is expected to reach 75 years in mid-twenty-first century (Riley 2001). Perhaps a dedicated decision analyst would claim that this is due to decision analysis itself, to the education of people in manipulating the apparatus of probability theory. That would, of course be outright wrong, for it would assume



that choices guiding the progress of medicine, biology, physics, manufacturing, and so on – which all contributed to the increased life expectancy one way or another – have been taken by following the tenets of decision analysis. It is *prima facie* evidence that very few are educated in the seeming sophistication of decision analysis. If *prima facie* evidence is not good enough, the skeptic should here consider Paul Slovic and Amos Tversky's experimental results (Slovic and Tversky 1974): subjects made choices and were then explained in an authoritative manner that they should not have chosen as they have; after reflecting on their choices given these arguments, most of them did not change their preferences over the choices they were initially offered, hence staying more content with choices that violate axioms of expected utility theory and those of decision analysis.

Besides these platitudes, there is another issue. When the decision analyst believes that probability theory is the appropriate instrument to handle uncertainty, he is in fact making a choice under uncertainty: what is uncertain is that the said instrument is indeed the one adequate for handling uncertainty in decision-making. To choose this instrument, and being a decision analyst, he must estimate the probability that probability theory is an adequate instrument in the face of uncertainty. How will he find that estimate in practice? Did some primordial decision analyst have the insight that probability is indeed the right tool? Decision analysis was advanced as the applied branch of economic decision theory of the 1960s, when decision theory was essentially expected utility theory, and thus axioms of perfectly rational choice were inevitable. There was apparently not much debate whether probability (and utility) should be used – they simply were there for the taking, and no other alternative was as accessible. It must have seemed *prima facie* evidence at the time that people should in fact use probability to handle uncertainty. This is today still a proposition too often taken at face value.

Perhaps there are those who do manage to convince themselves that they choose more rigorously after, so to speak, becoming more rational – at least according to what rational is by the criteria of the neoclassical economic model of man – by learning the ways of probability estimation promoted in decision analysis. The systematic error they thereby make is to leave unexplored other paths to handling uncertainty. Namely, they fail to consider the consequences of the possibility that *individuals may have difficulties with probabilities not because they cannot handle uncertainty, but because probability is not an appropriate instrument to interpret the decisions they actually take, within their actual decision contexts*. Now, the argument here certainly is not intended to discredit probability theory, but simply to consider the consequences of removing probability from the central role it still has in conceptualizing how individuals may be choosing. To do so is to avoid imposing rules onto decision-makers that they seem to systematically violate in practice.

Taken alone, the mathematical framework of probability theory is certainly a solid construction that admits various uses. In order to admit its various applications, the framework itself does not posit a relationship between the signs used in the framework and, so to speak, the real world and the people in it. To put this bluntly, the mathematical framework of probability theory – as in Kolmogorov's formulation for instance (Kolmogorov 1956) – does not say what probability is, beyond it being a concept that satisfies some axioms. It is then clear that if a model of choice uses

probability theory, then that model of choice *must* explain why it does so; such an explanation must establish precisely that relationship, which remains out of the mathematical framework itself and lets us meaningfully interpret the results that the mathematics provide in the specific context of application. It is only such an explanation that justifies the use of probability theory in descriptive and prescriptive models of choice. To be blunt again: the model of choice should explain what the term *probability* means, i.e., what that term refers to. It is only if an individual accepts that referent, that a model of choice suggests for the term *probability*, that this individual can accept the use of probability theory within that model of choice. The ensuing question is what referent does probability obtain in perfect, bounded rational models of choice, and decision analysis? In other words, what is the referent of the concept of probability, present in the ontologies of all these models?

### Probability and Its Antireferent

The term *probability* refers to a concept, which as any concept groups some properties. For something to be called the *probability of an event A*, it should satisfy the following axioms, synthesized by Andrey Kolmogorov, a twentieth century mathematician (Kolmogorov 1956, p. 2)<sup>1</sup>:

1. For a set  $E$  and a set  $\mathfrak{F}$  of subsets of  $E$ ,  $\mathfrak{F}$  is a field of sets. In other words,  $\mathfrak{F}$  is a nonempty subset of the power set of  $E$ , such that these hold:
  - (a)  $\emptyset \in \mathfrak{F}$
  - (b) If  $A, B \in \mathfrak{F}$  then  $A \cup B \in \mathfrak{F}$
  - (c) If  $A \in \mathfrak{F}$  then the complement  $\bar{A} \in \mathfrak{F}$
2.  $\mathfrak{F}$  contains the set  $E$ .
3. To each set  $A$  from  $\mathfrak{F}$  is assigned a nonnegative real number  $P(A)$ . This number  $P(A)$  is called *the probability of the event A*.
4.  $P(E) = 1$ .
5. If  $A$  and  $B$  have no element in common (i.e., are disjoint), then  $P(A \cup B) = P(A) + P(B)$ .

It should be clear that these axioms make reference neither to the term *uncertainty*, nor its synonyms. Their only result is to make explicit the properties that a real number must have to be called *probability of an event A*. In other, blunt words, we could replace above the term *probability of an event A* with, say, *greenness of the event A*, and the axioms will of course lose none of their relevance. This simply echoes Kolmogorov's introductory remarks to his treatment of probability:

[...] after we have defined the elements to be studied and their basic relations, and have stated the axioms by which these relations are to be governed, all further exposition must

---

<sup>1</sup> For a historical account of where these axioms come from, Glenn Shafer and Vladimir Vovk's discussion is relevant (Shafer and Vovk 2006).

be based exclusively on these axioms, independent of the usual concrete meaning of these elements and their relations [...] the concept of a *field of probabilities* is defined as a system of sets which satisfies certain conditions. What the elements of this set represent is of no importance in the purely mathematical development of the theory of probability [...] Every axiomatic (abstract) theory admits, as is well known, of an unlimited number of concrete interpretations besides those from which it was derived. (Kolmogorov 1956, p. 1)

The axioms consequently give the essential properties of the concept called *probability of an event A*, while leaving some freedom of interpretation. The consequence is that if we intend to use the term probability in a particular setting (e.g., when designing a model of rational choice), then we must establish additional reference relations, this time to objects other than (and in addition to) the properties defined by the axioms. In other words, probability is a concept that has to be tailored to particular settings, even though its basic properties remain those in the axioms. Because probability is a term having an antireferent, the question relevant here is what else has been added to its basic properties when that term is used in a model of choice?

Kolmogorov's interpretation of probability, what he called "the relation to experimental data" (Kolmogorov 1956, Sect. 2) is grounded in two principles outside of the axiomatic definition. Let  $\mathfrak{C}$  be some conditions, which can be repeated any number of times (e.g., a single tossing of a coin):

1. *Principle A*: "One can be practically certain that if the complex of conditions  $\mathfrak{C}$  is repeated a large number of times,  $n$ , the if  $m$  be the number of occurrences of event  $A$ , the ratio  $m/n$  will differ very slightly from  $P(A)$ ."
2. *Principle B*: "If  $P(A)$  is very small, one can be practically certain that when conditions  $\mathfrak{C}$  are realized only once, the even  $A$  would not occur at all."

Principle A reflects the relation between probability and frequency of occurrence, when exact same conditions are repeated a great number of times. It thus reflects the law of large numbers, that is, "[a] general principle according to which under certain very general conditions the simultaneous action of random factors leads to a result which is practically nonrandom. That the frequency of occurrence of a random event tends to become equal to its probability as the number of trials increases (a phenomenon which was probably first noted for games of chance) may serve as the first example of this principle" (Hazewinkel 2002). Historically the original form of the law of large numbers was Jakob Bernoulli's theorem, stating that, in a sequence of independent trials, in each of which the probability of occurrence of an event  $A$  has the same value  $p$ , with  $0 < p < 1$ , the relationship  $P(|\mu_n/n - p| > e) \rightarrow 0$  is valid for any  $e > 0$  if  $n \rightarrow \infty$ .  $\mu_n$  is the number of times that  $A$  occurred,  $n$  is the total number of trials (a trial, in Kolmogorov's terminology, equates to a situation, in which the conditions  $\mathfrak{C}$  hold). Principle B echoes what has been usually referred to as *the Cournot principle*, after Antoine Augustin Cournot, a nineteenth century mathematician and economist. Not only does it state that an event having very small probability will not happen, but that only because of this does probability theory have empirical meaning (i.e., relevance beyond the mathematical framework that its axioms allow) (Shafer and Vovk 2006).

We will not discuss why it was that Principle B was believed to be the only link between probability and, so to speak, the real world. What is instead interesting at present is another idea that followed from there, according to which there exist *objective probabilities* (Shafer 2006). There are roughly two ways to understand what objective probability refers to beyond the properties coming out of the axioms cited above. The *objective probability of an event A* may be (1) a quality inherent in the time series, in which the occurrence of *A* is observed, or (2) a quality inherent to the event *A*. In the first view (Shafer 2006), An *objective probability of an event A* is then defined as a quality inherent to a time series. A time series is a sequence of some number of trials, that is, of some number of repetitions of identical conditions. The quality to measure is *the event A taking place* in the given time series, or equivalently, that the time series “exhibits” the occurrence of *A*. In the other view (Szabó 2007), *objective probability of an event A* is the quality inherent in the event itself, not the time series in which it is observed. Whatever the view one takes, measuring probability will involve both the event and the time series, in which it occurs. Precisely because it cannot occur outside some specific conditions, the probability of an event cannot be measured on the event itself regardless of the time series. If it were possible to measure the probability of an event regardless of the time series, then  $n = 1$  (i.e., a single trial) would be acceptable and the law of large numbers would no longer apply. Once one assumes that there are objective probabilities, the next step is to measure them, just as, say, length and weight are measured. Principle A plays a key role in such measurement: “In a series of trials repeated a large number of times under identical conditions, each of the possible events happens with a (relative) frequency that gradually equals its probability. The approximation usually improves with the number of trials” (Castelnuovo 1919 cited in Shafer 2006, p. 5) Frequency is then a measure of probability. In this view, the term (*objective*) *probability of event A* no longer has an antireferent, but a definite referent, namely the quality of the time series, in which event *A* is observed to occur (or, if still preferred, the quality of the event *A* itself).

If objective probability is a quality of a time series, we cannot measure it if we cannot repeat the conditions holding at each point of that time series. This seems acceptable for well-defined experiments, although repetition under exact same conditions was already seen as unrealistic since Bernoulli (Shafer 2006). Possibility to repeat identical conditions evidently poses serious problems outside experiments, so that rare or unique events cannot be repeated (e.g., two planes hit the twin towers in New York, on 11 September 2001; the price of a stock hitting the same level for the exact same “reasons” as when it was at that level in the past; and so on).

Another view of probability – dominant in twentieth century economics – was developed from the 1930s onwards, by Frank P. Ramsey (Ramsey 1931), a mathematician, and Bruno de Finetti, a probabilist and statistician (de Finetti 1974). This approach, which speaks of a *subjective probability of an event A*, rejects that probability is an essential quality of an event or its time series. Rather, it argues, probability is in the eye of the beholder, a quality that the individual experiencing the event attributes to/sees in the event and/or its time series. If we ask in this view where probabilities come from, this is a mainstream answer:

probabilities are ‘in the mind’ — the subject’s, say, yours. If you say the probability of rain is 70 percent you are reporting that, all things considered, you would bet on rain at odds of 7:3, thinking of longer or shorter odds as giving an unmerited advantage to one side or the other. (Jeffrey 2004)

Subjective probability here refers to something called a “degree of belief.” Given some evidence, other kinds information, and so on, the individual does not see the occurrence of an event — e.g., “Earth will be visited by aliens from Mars.” — as a binary question, with this event being either certain (it will happen) or impossible (it will not/never happen). Rather, he is assumed to have a degree of belief in the occurrence of the event. According to this story, his subjective probability will refer to his degree of belief, e.g., 0.001 that aliens from Mars will indeed visit Earth. To measure degree of belief, de Finetti suggested a betting scheme:

Suppose we are interested in someone’s subjective probability that heads will turn up on the next toss of a given coin. De Finetti’s idea is as follows. We tell the subject: ‘We determine the stake on heads to be \$10. You may choose the stake on tails to be whatever you like, say  $Y$ . We then choose which side of the wager we take, and you have to play against us. I.e., we choose one of the two wagers: (1) you win \$10 from us if heads, you pay  $Y$  if tails, (2) you pay us \$10 if heads, you win  $Y$  from us if tails. De Finetti reasons that the subject must choose  $Y$  such that his expectation on both wagers is equal to zero. (Cooke 2004, p. 24)

If the individual chooses along the lines of expected utility theory, de Finetti argued that the subjective probability is  $Y/(Y + 10)$  in the case described above.

What probability may refer to, i.e., what is its appropriate antireferent, is a matter of a debate lasting for at least the last three centuries (Shafer and Volk 2001). In the economic models of man, grounded in one way or another in the axioms of perfect rational choice, the probabilities are of the subjective kind, in the sense of de Finetti (Shafer 2006). This is not to say that the notion of subjective probability somehow removes and replaces objective probability, merely that objective probability did make assumptions which could not be carried over to economics. It is clear now, however, that subjective probability has many problems itself (Cooke 2004; Shafer 2006; Hjek 2009; Szabó 2009).

Let us now return to our earlier question: what referent does probability obtain in perfect, bounded rational models of choice, and decision analysis? Its referent is something called a degree of belief, an individual’s quantification of his conviction that an event will occur. Since this referent is “in the mind,” it is an antireferent, tied to private experience. In terms of conceptual spaces, an estimate of subjective probability is a numerical value that obeys the axioms of probability theory, and that refers to a range in the individual’s conceptual space for a degree of belief or conviction. It is interesting to observe that subjective probability therefore postulates that the individual, the human decision-maker has something akin to a conceptual space for degree of belief or conviction and he is capable of attaching a numerical value to ranges in that conceptual space. Subjective probability is thus not the quality inherent to the event or the time series, in which the event occurs (as objective probability purports to be), but the quality inherent to the individual. To accept then, the normative mission of decision analysis requires that we also accept the assumption that individuals do have that inherent quality, which is to evaluate own degree

of belief, but that they are not terribly good at using it (as empirical evidence tends to indicate).

## **Taking No Sides**

If we leave aside objective probability as something unsuitable for a model of choice, the subjective reading of probability still cannot negate the observation that if we are interested in advising a decision-maker, we are not going to advise them on probability values. Even the subjective probability of decision-making models is recognized as a summary of an individual's information, evidence, and whatnot:

Your 'subjective' probability is not something fetched out of the sky on a whim; it is your actual judgment, normally representing what you think your judgment should be, in view of your information to date and of your sense of other people's information, even if you do not regard it as a judgment that everyone must share on pain of being wrong in one sense or another. (Jeffrey 2004)

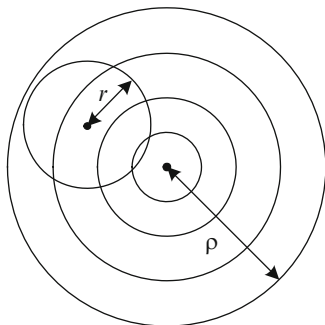
To the extent then that a probability estimate is a summary of the information available to the decision-maker, it is produced by his reasoning on/about that information. To conclude a probability value, or rather, an estimate of probability, he uses these premises that are available to him, and a method of inference, to move from the premises and reach his conclusion. As long as there is such a step in the estimation of a (objective or subjective) probability, accepting advice on simply another probability value is much too easy, for it would make irrelevant the premises and inference invested in coming up with the initial probability estimate in the first place. It follows that advice will need to target the premises and/or the method of inference that the decision-maker applied to arrive at the probability estimate.

If probability does not merit its central role in a model of choice relevant to the analysis of advice, then should we expel it altogether? The positive answer is not acceptable for anyone who accepts either the objective or the subjective reading of probability. Those favorable to objective probability will argue that there are decision settings where frequency of occurrence of an event reflects the objective probability of that event in a random and well-defined experiment. Those preferring the subjective reading of probability will argue that we can still do some reasoning with subjective probability estimates, and that the results of that reasoning can help grasp something about the decision problem. Middle ground is not to choose any of these two readings, but to see any probability estimate as just another kind of information that can be accounted for when the decision-maker reflects on the decision problem. That middle ground will make no commitment to either a subjective or objective probability, but remain silent on that matter. For that silence to be acceptable, we need an ontology of choice, in which probability estimates do not play the prominent role they have in models of perfect or bounded rational choice, and decision analysis. It is only from such a model of choice that we can then obtain an ontology of advice, which can have advice that targets the rationale behind whatever estimate of probability may be used.

Interestingly and conveniently enough – as will become clear in the rest of this chapter – our ontology of choice will also allow a commitment to a third option with regards to probability. That option says explicitly that the term *probability* has an antireferent, that the properties/objects it refers to will differ depending on the context of use. There is no commitment that probability is either a quality of an event, or of a time series, or of the individual. Instead, as László Szabó, a physicist argued “[w]henever we use the term ‘probability’ in scientific discourse, its meaning varies from context to context: it means different dimensionless  $[0, 1]$ -valued physical quantities, or more precisely, different dimensionless normalized measures composed by different physical quantities in the various specific situations. Moreover, these context-dependent meanings reduce the concept of ‘probability’ to ordinary physical quantities of empirical meanings, like relative frequency on a finite sample” (Szabó 2009, p. 6). In other words, when the term probability is used, it is used as a shorthand, a replacement for some relation between measures on observable phenomena. Probability of the event then refers to those relations only and not to some quality inherent to the event, the time series in which the event is observed, or the individual doing the observation. Consider the example in Fig. 5.1. The only way to read this equality is that the sign  $P(A)$ , which we tend to call by convention the probability of the event  $A$ , is defined as  $\pi r^2 / \pi \rho^2$ . The signs in  $\pi r^2 / \pi \rho^2$  are known physical quantities, amenable to measurement.  $P(A)$  is thus here only a shorthand for  $\pi r^2 / \pi \rho^2$ , and refers only to  $\pi r^2 / \pi \rho^2$ , and not to some quality of the event, of a time series of trials, or of the individual observing the shooting of the gun at the target.

### 5.3.2 Utility Intolerance

The standard theory of decision-making under risk and uncertainty, and certainly the dominant choice paradigm since the Second World War is expected utility



**Fig. 5.1** “A gun is hinged in such a way that it can fire uniformly at a round target area, radius  $\rho$ , with an inflated balloon, radius  $r$ , attached to the front of the target. What is the probability that the balloon will be burst (event  $A$ )? The physicist’s standard answer to this question is the following:  $P(A) = \pi r^2 / \pi \rho^2$ .” (Szabó 2007, p. 629)



theory. The conceptualizations for perfectly rational choice and boundedly rational choice, along with those underlying decision analysis cover the basic formulation of expected utility theory.

Variants of the expected utility conceptualization differ in (1) how utility is measured, (2) what kind of probability transformations are allowed, and (3) how the outcomes are measured. It is not necessary to go beyond Paul Schoemaker's review of these variants (Schoemaker 1982) to see that the other approaches do not contain concepts or relationships that are not transformations of those outlined earlier for perfectly rational choice. Generalizations of expected utility theory have been proposed to address empirical evidence showing that theory to be descriptively misleading. Chris Starmer's review (Starmer 2000) of these various generalizations, each aiming for the so-called nonexpected utility theory, indicates that some conceptual extensions have been added to the basic conceptualization. Among these is John Quiggin's proposal (Quiggin 1982) for rank-dependent expected utility theory. There, the aim is to maximize a weighted sum of utilities; the weights add up to one. It differs from expected utility theory in that the weight of an outcome depends on its probability and its rank relative to other outcomes. The concept of weight of an outcome is thus added to the core conceptualization, allowing it to make explicit that decision-makers subjectively "distort" objective probabilities. Prominent among procedural proposals (i.e., those that consider the sequence in decision-making activities) to nonconventional expected utility is Daniel Kahneman and Amos Tversky's prospect theory (Kahneman and Tversky 1979), which articulates the basic conceptualization of expected utility theory, while adding two distinctive points: (1) there is a pre-decision phase, in which some of the available alternatives are "edited" via various decision heuristics (mainly to obtain simpler representations of the alternatives), and (2) outcomes are interpreted as gains and losses relative to a reference point. Loomes and Sugden's regret theory (Loomes and Sugden 1982) argues that an individual considers alternatives pairwise, where the utility of the consequences of an act depends also on the consequences anticipated for the other considered act. Hence, what is and what might have been together determine the utility associated with each of the choices. The modified measure of utility then enters the standard expected utility conceptualization. It has the benefit of allowing nontransitive preferences.

Reproaches to the term *utility* are not unlike those discussed for the term *probability*. If there is a definite referent for *utility*, then it is defined intensionally via the axioms of expected utility theory (von Neumann and Morgenstern 1956), but also via the properties chosen outside the mathematical framework itself. The question becomes what utility is, a quality of the individual, of the anticipated outcome of a choice, or something else? A prominent interpretation, still influential in one form or another comes from Jeremy Bentham, an eighteenth and nineteenth century philosopher, who asks rhetorically "To what shall the character of utility be ascribed, if not to that which is a source of pleasure?" (Bentham 1825, Book III, Chap. I) His own conception is this:



By utility is meant that property in any object, whereby it tends to produce benefit, advantage, pleasure, good, or happiness, (all this in the present case comes to the same thing) or (what comes again to the same thing) to prevent the happening of mischief, pain, evil, or unhappiness, then the happiness of the community: if a particular individual then the happiness of that individual. [...] An action then may be said to be conformable to the principle of utility, or, for shortness sake, to utility, (meaning with respect to community at large) when the tendency it has to augment the happiness of the community is greater than any it has to diminish it. (Bentham 1823, pp. 3, 5)

The basic idea of utilitarianism is this view, according to which utility refers to a quality of the object to serve for the increase in the happiness of all, or to the prevention of events that would reduce happiness. Utilitarianism formulates moral directives, and as such has ambitions that go much farther than the axiomatic definition of utility in expected utility theory. It is only by adding something more to the axioms, another referent which involves notions of happiness and pain that such ambitions can be pursued. Amartya Sen, a philosopher and economist, argued that one such directive, *outcome utilitarianism* is basic to any variant of utilitarianism: “any state of affairs  $x$  is at least as good as an alternative state of affairs  $y$  if and only if the sum total of individual utilities in  $x$  is at least as large as the sum of individual utilities in  $y$ .” (Sen 1979b, p. 464) What Sen calls the utilitarian moral structure, a morality based on utilitarianism, combines the said principle of outcome utilitarianism with some way of relating outcomes to actions, for it is actions that need be taken toward the realization of the outcomes.

The use of outcome utilitarianism, its application toward policy making requires a tool, a framework of thought, which economics provides as one variant or another of expected utility theory. Utility and utilitarianism meet in the very definition of utility, part of the definition being provided via axioms within the mathematical framework itself, the other coming from pleasure or happiness as the referent of utility. Utilitarianist policy making seeks to identify the actions that would lead to the optimal outcome in terms of aggregate utility of individuals. It needs to assume a utility function identical to all individuals, a function that relates outcomes to utility levels, and then maximizes the aggregate of individual utilities. Thus the classical question of social choice theory ensues, namely whether interpersonal comparison of utility is a plausible idea. If so, the policy maker would estimate the individual utilities or, rather the aggregate utility of outcomes, then design policy which is expected to realize the outcome with the maximal aggregate value of individual utilities. To follow Bentham’s lead is to assume that this is a plausible idea, namely, that people can compare one another’s evaluations of utility: in a sense, there is a utility scale that is common to all, and the differences between the points on that scale are experienced in the same manner by everyone. Such a view gives legitimacy to the use of utility estimates to the design of policies, on, say, how to improve the conditions of the poor. Lionel Robbins went to the other extreme in his critique, arguing that it is impossible to evaluate how another would experience the consequences – in terms of happiness or pain – of some anticipated outcome. Kenneth Arrow later offered his impossibility theorem (Arrow 1950), stating that there is no

way, given each individual's complete preference orders<sup>2</sup> over all alternative outcomes, to choose an outcome for the collective that will satisfy all of the following three conditions together (Sen 1979a):

1. If every individual prefers  $x$  to  $y$ , then  $x$  is socially better (i.e., better for all individuals) than  $y$ .
2. There should be no dictator, that is, no individual, such that if he strictly prefers  $x$  to  $y$ , then  $x$  is immediately regarded as socially better than  $y$ .
3. The social preference or indifference between  $x$  and  $y$  should depend only on each individual's preference or indifference between  $x$  and  $y$ .

Sen argued that an individual's utility evaluation is malleable, and thus not a proper starting point for the interpersonal evaluation of utility values, the comparison of how well off different individuals are in relation to others: "A hopeless destitute with much poverty, or a downtrodden laborer living under exploitative economic arrangements, or a subjugated housewife in a society with entrenched gender inequality, or a terrorized citizen under brutal authoritarianism, may come to terms with her deprivation. he may take whatever pleasure he can from small achievements, and adjust her desires to take note of feasibility (thereby helping the fulfillment of her adjusted desires). But her success in such adjustment would not make her deprivation go away." (Sen 1999, p. 358) Differences in the utility functions of individuals will thus tend to increase inequality when policy makers wish to maximize total utility. It thus becomes unimportant whether there is some universal utility scale, manifested necessarily at each individual and allowing everyone to meaningfully compare one's utility estimate to anyone else's. It is more relevant to ask how something akin to utility should be measured otherwise: Sen thus speaks of a person's advantage, measured by, say, a person's access to "general-purpose resources that are useful for anyone to have no matter what her exact objectives are" (Sen 1999). These may include access to food, water and electricity, medical care, and so on.

The moral of this story, one interesting in the discussion of advice, is that the move from Benthamian utility to Sen's broad understanding of an individual's advantage in terms of measurable quantities (e.g., income, access to food, medical ) parallels the move from the objective and subjective conceptions of probability, to a conception in which a probability value is a shorthand, a replacement for some relation between measures on observable phenomena. If utility is used in a model of choice, it is merely a tool to summarize some other information. An intolerance for utility follows when we get interested in advice. While utility remains in models of perfect or bounded rational choice, it hardly obtains a Benthamian interpretation; as Kenneth Binmore, an economist observes:

[F]ar from maintaining that our brains are little machines for generating utility, the modern theory of utility makes a virtue of assuming *nothing whatever* about what causes our

---

<sup>2</sup> The utility function can be represented as a total preference order, so that if the utility of  $x$  is higher than that of  $y$ , then  $x$  is strictly preferred over  $y$ ; if the utility of  $x$  equals that of  $y$ , then the individual is said to be indifferent between  $x$  and  $y$ .

behavior. [...] The modern theory of utility therefore abandons any attempt to explain *why* people behave as they do. Instead of an explanatory theory, we have to be content with a descriptive theory, which can do no more than say that a person will be acting inconsistently if he or he did such-and-such in the past, but now plan to do so-and-so in the future. Such a theory is rooted in observed behavior. It is therefore called a theory of ‘revealed preference’, because the data we use in determining what people want is not what they say they want — or what paternalists say they ought to want — but our observations of what they actually choose when given the opportunity. (Binmore 2009, Binmore’s emphasis)

Analysis of advice cannot take this same stance, for advice is given precisely to influence the reasons for choosing one alternative over another: it works at the level of the *why*, before the choice is observed. To make matters worse, there are good reasons to believe that expected utility theory not only fails to be explanatory, but it also fails to be descriptive. Paul Schoemaker’s synthesis of that failure points to three problems (Schoemaker 1982): (1) people cannot be as comprehensive when finding and organizing information about the decision problem; (2) people neither estimate probabilities nor use them as the theory assumes; and (3) the theory poorly predicts choice behavior in laboratory settings. Normative use is also questionable, for empirically observed biases in decision-making “may be so basic that they render the normative theory inoperational” (Schoemaker 1982, p. 554), while the persistence of these biases further increases doubt about its normative relevance.

Conclusions reached here for utility and its estimates parallel the conclusion on probability and its estimates: advice on utility estimates remains very much uninteresting, utility estimates are summaries of something else, and it is that other which needs to be looked into in order to offer an ontology of advice. The intolerance to utility cannot be avoided, when analysis of advice is a matter of interest, and the aim is to identify kinds of advice.

At the turn of the twentieth century Vilfredo Pareto, an Italian economist argued that:

In reality and in the most general way, pure economic equations simply express the fact of a choice, and can be obtained independently of the notion of pleasure and pain. This is the most general point of view and also the most rigorous. [...] For us, it is sufficient to note the fact of individual choice, without investigating the psychological or meta-physical implications of such a choice. [...] We do not inquire into the causes of men’s actions: the observation of the fact itself is sufficient. [...] Pure economic equations and their consequences exist unchanged whether we start from the consideration of pleasure as a quantity, or we limit our investigation [...] exclusively to the fact of choice. (Pareto 1900, pp. 221–224) cited in Marchionatti and Gambino 1997, pp. 1335–1336

The stance of the advisor, an individual intolerant to substitutes, is then resolutely different from that which is allowed to, and favored by the economist.

## 5.4 Decision Information

A different model of choice is needed to get to an ontology of advice. The preceding section concluded that no model of choice will do as long as it is tolerant to substitutes, i.e., if probability and utility estimates in particular have a central role. Advice

on probability and utility estimates itself is of little interest. While this is not to say that it cannot be given or taken, it is to claim that there should also be advice on the content (i.e., what information is used in reflection about the decision situation) and process (i.e., how reflection proceeds) of the rationale that are said to be reflected in some given probability or utility value.

Judging from the perfect and bounded rational models of choice, and the others mentioned up to this point, there are three parts in a model of choice:

1. *Decision information ontology* categorizes inputs that the decision-maker takes into account when choosing. In decision analysis for example, the categories of inputs are *alternatives*, *objectives*, *attributes*, *probabilistic outcome function*, and *utility function*.
2. *Decision problem formulation* explains how the instances of a decision information ontology, the decision information come together to give a picture of the decision situation. By “together” is meant that relations are established between decision information: e.g., a probabilistic outcome function relates alternatives to outcomes.
3. *Decision Rules* define the properties which the decision-maker should perceive in a solution to the decision problem. Taking decision analysis again as an example, the decision rule there says that a solution should have two properties: (1) be a member of the set of alternatives (i.e., an instance of the concept *alternative*), and (2) give maximal expected utility compared to all other instances of *alternative* in the same decision problem.

The aim of this section is to give a different model of the decision-maker, and to use that model to develop an ontology of decision information different than those discussed earlier. Only then can an ontology of advice be offered, and later the *advisor's problem*, that is, the basic problem that any advisor needs to resolve when designing advice.

### 5.4.1 *Revealed Intentional States*

It is while moving toward the definition of *advice* in Chap. 3 that the critical role of communication was noted. Advice was indeed defined in relation to communication. The picture of communication that was drawn is one of speech acts. Communication is action, its components being speech acts, kinds of action that the speaker uses to influence the hearer's beliefs, desires, intentions, and so on.

The fundamental way in which, say, assertive speech acts differ from directive speech acts lies in that the two convey two different *Intentional states* of the speaker. If a speaker *S* uses an assertive speech act *a*, then the hearer would extract at least two pieces of information from *a*. Since a speech act incorporates the modus and dictum, mode, and content, the hearer can distinguish these two. The assertive speech act *a* can then be rewritten – by the hearer – as *S believes that p*, where *p* is the content of the speech act, and *believes*, or rather, *belief* is the intentional

state that the hearer believes the speaker is in. The hearer establishes the intentional state from the modus of the speech act. If  $S$  uses a directive speech act  $e$ , then the hearer can rephrase  $e$  as  $S$  desires that  $p$ , where the propositional attitude is that referred to by the term *desire*. The more general idea is that any speech act can be rewritten as  $S$ - $m$ -that- $p$ , where  $m$  refers to an Intentional state that the hearer infers from the modus of the speech act, and  $p$  is a sentence standing for the content of the speech act.

The view of communication through speech acts subsumes a model of man, or more specifically, a model of the mind. In that view, an intentional state is a specific kind of mental states. In an essay on *The Nature of Intentional States* (Searle 1983) John Searle argues that not all mental states can be called intentional states. The distinction lies, according to him in how the states are reported, spoken of: "If I tell you I have a belief or a desire, it always makes sense for you to ask, 'What is it exactly that you believe?' or 'What is it that you desire?'; and it won't do for me to say, 'Oh I just have a belief and a desire without believing anything or desiring anything.' My beliefs and desires are always about something. But my nervousness and undirected anxiety need not in that way be *about* anything. [...] Here are a few examples of states that can be Intentional states: belief, fear, hope, desire, love, hate, aversion, liking, disliking, doubting, wondering whether, joy, elation, [...]" (Searle 1983, pp. 1–4).

The rough model of the mind is then this: there are intentional states, they have a psychological mode and a content, namely propositions referring to objects. Communication results from the intention to convey intentional states, that is, contents plus psychological modes. How is this related to the model of the decision-maker, or rather, how can this model of mind be a starting point for an ontology of decision information?

## Where Concepts May Come From

To answer the question just posed, start with another one: what are the criteria used to determine what goes in and what stays out of an ontology of decision information? It is precisely such criteria that determine what concepts are found in an ontology. Before concepts are chosen, ontological commitments are made, and it is the ontological commitments that explain and determine the criteria that a concept must satisfy to enter an ontology. Needless to repeat that finding and justifying the use of some such criteria over others is a difficult task with debatable results, as was already discussed when we gave the ontological commitments behind the definition of advice, in Chap. 3.

What are, or could have been then the criteria that have led to introduce, say, the concepts of (*behavioral*) *alternative*, (*decision*) *outcome*, *payoff function*, and *probabilistic outcome function* into the ontology of decision information in models of perfect and bounded rational choice? This may seem a strange question to someone acquainted with the cited models and their merits in economics, in terms of the insights they have brought about, be it through their use in theoretical or empirical

developments, or the results gained through criticism thereof. Invaluable as they are, it is not because the question is strange that it should not be pursued. It is moreover a question that does not admit an easy answer, for it is unclear precisely who and when first started using notions of *alternative*, *outcome*, and so on, in thinking about choices, and if they explained why they did introduce precisely these concepts and not others.

Consider von Neumann and Morgenstern's *Theory of Games and Economic Behavior* (von Neumann and Morgenstern 1956), which introduced the axioms of perfectly rational choice, and set the bases for the theory of games. A reading of their first chapter – which lays out their assumptions about the method of economics, the problem of what rational behavior may be, the concept of utility, and on what a solution of a game is – gives the impression that the key insight from which the entire development stems is the similarity they see between the behavior of actors involved in economic transactions and players of a game:

We think that this similarity is very essential; indeed, that it is more than that. For economic and social problems the games fulfill — or should fulfill — the same function which various geometrico-mathematical models have successfully performed in the physical sciences. Such models are theoretical constructs with a precise, exhaustive and not too complicated definition; and they must be similar to reality in those respects which are essential in the investigation at hand. [...] It is clear that if a model of economic activities is constructed according to these principles, the description of a game results. (von Neumann and Morgenstern 1956, Sect. 4.1.3)

Their basic assumption is that economic interactions are alike interactions in games. The ontological commitment that ensues is that the concepts for a theory of games, and thus for a model of choice of the participants in the game should reflect the commonsense understanding of common elements in games (e.g., chess, tennis, and so on):

First, one must distinguish between the abstract concept of a *game*, and the individual *plays* of that game. The *game* is simply the totality of the rules which describe it. Every particular instance at which the game is played — in a particular way — from beginning to end, is a *play*. Second, the corresponding distinction should be made for the moves, which are the component elements of the game. A move is the occasion of a choice between various alternatives, to be made either by one of the players, or by some device subject to chance, under conditions precisely prescribed by the rules of the game. The *move* is nothing but this abstract 'occasion,' with the attendant details of description — i.e. a component of the game. The specific alternative chosen in a concrete instance — i.e. in a concrete *play* — is the *choice*. Thus the moves are related to the choices in the same way as the game is to the play. The game consists of a sequence of moves, and the play of a sequence of choices. Finally, the *rules* of the game should not be confused with the *strategies* of the players. [...] Each player selects his strategy — i.e. the general principles governing his choices — freely. [...] The rules of the game, however, are absolute commands. If they are ever infringed, then the whole transaction by definition ceases to be the game described by those rules. (von Neumann and Morgenstern 1956, Sect. 6.1; von Neumann and Morgenstern's emphasis)

The ontology of game theory is reflected by, or itself reflects and builds upon the ontology of decision information from expected utility theory. The answer to the earlier question is then this in game theory: the criteria that a concept should satisfy to remain in the ontology of game theory are determined by the intuitive

conception of common games, such as chess. A concept remains in the ontology if it captures some property perceived in various common games, and as long as it is not synonymous and does not overlap with already adopted concepts. The model of the perfectly rational decision-maker thus is one of a player in a well-defined game, and all concepts (e.g., *alternative*, *payoff function*, etc.) arise out of *prima facie* evidence about what matters to players in common games.

## Communication Reveals Intentional States

The simple claim that *economic interactions can be thought of as common games* is the starting point for the development of the ontology of game theory. What is that starting point in our model of choice for the analysis of advice? It is the simple claim that *intentional states and actions are interdependent*, which is then the starting point in the following line of thought:

1. An individual's intentional states and his actions are interdependent. It is not the case that intentional states and actions are fully independent.<sup>3</sup>
2. Intentional states influence at least some actions; actions and their observed effects influence intentional states.<sup>4</sup>
3. At least some actions are performed because one has decided to act in the way that was observed while he was performing the chosen action. Stated otherwise, choice depends on intentional states.
4. Because choice depends on intentional states, the decision information ontology of a model of choice should include concepts that capture all kinds of intentional states. Some of the factors influencing choice would otherwise remain obscure.

The conclusion that *the decision information ontology of a model of choice should include concepts that capture all kinds of intentional states* results in a criterion for the inclusion of a concept in such an ontology. It also leads to the difficult question of *what kinds of intentional states there are?*

There seem to be many intentional states (Searle 1983; Crane 2001): to the extent that intentional states are something inferred by the hearer about the speaker, and thus inferred from the communication between these two, the multitude of intentional states is observable from the verbs that refer to psychological modes. Anything that can reasonably be put in place of *m* in *S-m-that-p* is an admissible sign for the psychological mode of an intentional state. Do some of these signs have same referents, are they just different signs for the same psychological modes, and

---

<sup>3</sup> It is indeed not controversial to say that one's actions depend on one's beliefs, desires, and so on (whatever these may actually be – see below): e.g., an individual may desire to climb the Mount Everest, but may not intend to do it because he believes that this is too difficult.

<sup>4</sup> Two remarks: (1) the claim is made only for some and not all actions, because it is unclear whether reflex actions (e.g., patellar reflex, or usually known as knee-jerk) have something to do with intentional states; and (2) if the actions and their observed effects did not influence intentional states, learning would not be possible.



thereby for same kinds of intentional states? Also, are some of them not combinations of others? Searle answers both questions positively (Searle 1983). He claims, for example, that “*S* is sorry that *p*” can be reduced to “*S* believes that *p* and desires that the opposite of *p*” (Searle 1983, p. 32). His positive answer is cautious, for it is unclear whether many other psychological modes could be reduced at all to only a combination of belief and desire – e.g., what is to be amused about something in terms of beliefs and desires?

This is the point where we should ask the question of *what is an intentional state*? An answer that would fit in ontology, as in metaphysics, will not be sought and in general is considered uninteresting for the present discussion. Reasons will be clearer later on, but the main point is that such a position is simply not needed to make the claims that are indeed offered in the rest of this and subsequent chapters.

The difficulty to establish just which psychological modes of intentional states are primitive, and the inability to estimate their number makes it necessary to commit to some further simplifications. This leads here to the following assertions that continue the line of thought given earlier:

#### 5. Communication (verbal and nonverbal) reveals intentional states.

How does this help? It says that *intentional states, whatever they are, are ascribed to the decision-maker by an observer, and that it is from the communication with the decision-maker that the observer decides which intentional states to ascribe to the decision-maker*. It may not seem apparent how the fifth assertion above and this claim are related, how to go from the former to the latter. Number 5 says that intentional states are revealed via communication, be it verbal or otherwise. Is there no other way to discover others’ intentional states? To the extent that true telepathy and mind-reading are elusive, communication remains the only plausible option. Communication certainly is not confined to verbal exchanges. It should rather be taken in a very wide sense. For example, if the decision-maker is not visibly engaged in the communication with the observer, something can still be communicated: suppose that the observer is in the audience at a tennis match, and the observed decision-maker is a player in that match – if the player changes how he is playing his opponent, say, by approaching the net more often, then this act does in itself communicate something, which the observer may interpret as the result of that player’s belief that his opponent is weaker when attacked from the net. It should be clear that in total absence of communication, the observer’s assumptions about the intentional states of the decision-maker are speculations coming out of thin air.

To say that communication reveals intentional states is to say that two questions can remain unanswered: (1) what intentional states actually are, in the sense of their place in some universal ontology, as in metaphysics; and (2) whether the intentional states conveyed by communication are effectively those that are “in the mind” of the observed decision-maker, or in other words, whether the intentional states that the observer ascribes to the observed decision-maker are those that the latter truly holds. As long as these two questions remain open, the term *intentional state* can only point to an antireferent, and *not* a definite referent.



If the phrase *intentional state* points to an antireferent, then this question must be answered: what properties *should* (as opposed to *does*) the antireferent of *intentional state* have? The properties it should have need to reflect the role that phrases such as *S-believes-that-p* have for the observer of *S*, when that observer aims to anticipate the behavior of *S*.

What role then, do Intentional states have in choice? Recall that the overall aim here is the analysis of advice: the advisor is normally not simultaneously the recipient of advice (i.e., one does not advise oneself, but someone else): there must be at least two people for advice to be exchanged. These two people play different roles at different times, either of the advisor or of the recipient. In each role, they will ascribe to the other person something that we will call, for all practical purposes, *intentional states*. What justifies this claim is the *prima facie* evidence that language incorporates means to speak about what one “believes” the other “believes,” “desires,” “fears,” and so on. It is language that permits the ascription of intentional states to happen, and there must be something to it – such elements of language would be used very little, if at all, if they did not perform some role in communication and the description of others’ behaviors. What these intentional states *are exactly* is secondary to what *purpose* the corresponding elements of language, the signs themselves have: when advice is being exchanged, the selection of what to advise will be influenced by the intentional states that the advisor has ascribed to the recipient. *This selection needs not assume what the intentional state is exactly, but it must assume – at least as far as advice-giving and taking is concerned – what may have led to this ascription in the first place, and what can be inferred about future behavior of the from that ascription.* A nonsmoker may advise the smoker to read a pamphlet on the negative health effects of smoking if he believes that the smoker believes that smoking is not harmful: in doing so, the nonsmoker-advisor has ascribed an intentional state to the smoker, say, the belief that smoking is not harmful.

The story of “*A* believes that *B* believes *X*” – i.e., the whole story of what it means to ascribe intentional states – is not that *A* somehow reads the mind of *B* and knows what is “in there” (if anything), but that when *A* says “I believe that *B* believes that *X*,” *A* is saying something along these lines: “I (i.e., *A*) anticipate that *B* will act in the same way that I would act if I were to ascribe to myself the beliefs, desires, and so on, that I ascribe to *B*.” Now, the question of how *A* will anticipate that *B* would act – e.g., as a perfectly rational decision-maker, as a decision analyst, as a psychopath, and so on – ends up depending not on *B*, but on the set of intentional states that *A* ascribes to *B* and the conclusions about what to do that *A* infers from these ascriptions.

Consider again the nonsmoker and the smoker, the former advising the latter to read a pamphlet on the harms of smoking. According to the picture painted above about the ascription of intentional states, the nonsmoker will give that advice to the smoker *only if* the conclusion of the former’s reflection leads him to conclude that the intentional states he had ascribed to the smoker may change as the result of advice, or if the very act of advice plays some role beyond the potential impact it could have on the immediately targeted recipient. In this latter case, the nonsmoker

advises the smoker to read a pamphlet not because this might affect the future behavior of the smoker, and thereby the nonsmoker's future ascriptions of intentional states to the smoker, but because the very act of giving this advice serves to influence the intentional states of those that can perceive the advice-giving scene: e.g., the nonsmoker desires to influence the intentional states that others ascribe to his (e.g., that he deems smoking undesirable), so that the advice he gives to the smoker is not only there for the smoker, but is there for those others who perceive the scene. Whoever's ascriptions of intentional states are targeted, the nonsmoker will need to do two things: (1) he will have to ascribe intentional states to the target of his advice, and from there (2) reflect on how he would act if he were to ascribe to himself these intentional states he has ascribed to his target. Only then does it make sense for the nonsmoker to dispense advice. Clearly, the best the advisor can do is speculate from one's own perspective.

This idea that to predict others' behaviors, the observer ascribes intentional states to them and reasons from there does not seem far from either the so-called theory-theory (TT) and simulation-theory (ST) accounts of how individuals ascribe to others some beliefs, desires, intentions, and the various other intentional states. Such ascriptions perform a role, which is the anticipation of others' behavior toward the adjustment of one's own. Both TT and ST are concerned then with what some call *mentalizing*, understood as a process by which an observer makes inferences about the intentional states of the individual or individuals he is observing. As to how mentalizing happens, psychology and neurology suggest that:

[m]any cues in different modalities can trigger the process of mentalizing as long as they originate from an agent. Agency can be perceived in other animals and even in moving objects (Heider and Simmel 1944), but the agents we are most interested in are our conspecifics. Their faces, in particular, are an important source of information about their inner states. For example, there is agreement about what a trustworthy person looks like even though this is an example of a prejudice with little basis in reality. Emotions, on the other hand, can be validly read from facial expressions, from voices, and from whole-body movements (Adolphs 2002). Desires, goals, and intentions can be read from eye gaze direction and body movements (Langton et al. 2000). Beliefs are computed by recognizing that knowledge depends on experience, so that someone may not know what we know because they have not seen what we have seen (Wimmer et al. 1988). Note that this example involves perspective taking, a vital aspect of successful mentalizing. Communicative intentions are perceived when someone calls our name or makes eye contact (Sperber and Wilson 1995). (Frith and Frith 2006, p. 531)

According to TT, the observer starts from assuming that the observed individual has some beliefs and desires. To predict how the observed individual will act, the observer makes some assumptions, such as that people seek to satisfy their desires, that they do not act in violation of their beliefs, and that people are unhappy when desires are not fulfilled. From there on, the observer assumes that the other will act in some specific way to satisfy own desires and not violate own beliefs. Where ST differs from TT is that ST assumes that the observer will also attempt to mimic, impersonate, replicate the beliefs, desires, intentions that he ascribed to the observed individual: the observer will reflect as if it was his who held the intentional states ascribed to the observed individual. he will predict how the observed individual will

behave by pretending that he holds the intentional states that he ascribes to that individual. TT will predict the behavior of the other not from how the observer may act if he held the intentional states that he ascribed to the other, but from how some general principles (e.g., that people aim to satisfy their own desires) that tell how people behave in general. The hybrid of TT and ST may be closer to its target as a theoretical model of how mentalizing happens, for there are compelling arguments that both these approaches are used when anticipating or explaining the behavior of others (Nichols and Stich 2003). For instance, when people systematically fail to correctly predict others' actions, this has been interpreted as indicating that they hold an inadequate theory of how others decide to act; in other words, they are anticipating others' behavior in a way that TT postulates (Apperly 2008). A compelling argument in favor of ST is that there are so-called mirror neurons in the brain, which activate not only when an individual performs some action, but also when he perceives another individual performing the same action:

The human brain is endowed with structures that are active both during the first- and third-person experience of actions and emotions. When we witness someone else's action, we activate a network of parietal and premotor areas that is also active while we perform similar actions. When we witness the disgusted facial expressions of someone else, we activate that part of our insula that is also active when we experience disgust. Thus, the understanding of basic aspects of social cognition depends on activation of neural structures normally involved in our own personally experienced actions or emotions. By means of this activation, a bridge is created between others and ourselves. (Gallese et al. 2004, p. 400)

Both the TT and ST are variants of what Daniel Dennett, a philosopher calls the *intentional strategy* for predicting the future behavior of the person:

Here is how it works: first you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have given its place in the world and its purpose. Then you figure out what desires it ought to have, on the same considerations, and finally you predict that this rational agent will act to further its goals in the light of its beliefs. A little practical reasoning from the chosen set of beliefs and desires will in many — but not all — instances yield a decision about what the agent *ought* to do; that is what you predict the agent *will* do. (Dennett 1990, Dennett's emphasis)

It is safe to say that the variety of both theoretical and empirical work on TT and ST make it hard to discard mentalizing, or intentional strategy as something entirely obscure and speculative at best.

Recall that our aim is to design an ontology of decision information, so what, then, do we commit to here toward that aim? What are the conclusions of the preceding discussion — do we build an ontology on the intentional strategy, ST, TT, or something else? None of these explicitly, and not because we disagree, but because there is in a sense too much in all of them. We are only interested for the moment in ontological commitments that precede the choices of concepts for a decision information ontology. The only thing we do need to commit to is that *intentional states are not in the mind of the observed individual, and neither in the eye, but in the language of the observer*. It is therefore unimportant for the design of the decision information ontology if these intentional states are something that can be found “in the mind” or in the wiring of the brain. That there is support for intentional states

beyond language and in the brain, as to some extent with mirror neurons is certainly relevant, but only insofar as it justifies the idea that others' behavior is anticipated through the ascription of intentional states and some reasoning thereon.

To say then that people do seem to reason about others' choices by ascribing intentional states to them, and that this position find support in philosophy, neurology, and psychology, leads to a nonobvious rewriting of all five assertions stated earlier. The crux of the change is that there is no need to understand what precisely an intentional state is, so to speak "within the mind" or whatever consideration of this sort. It is enough that intentional states refer only to combinations of assumptions and expectations that the observer/advisor makes about the behavior of the other/receiver. Whatever more one wishes to believe, e.g., some more or less elaborate idea of what mind may *really* be (as in metaphysics) is left for those interested enough in such an issue. It is in this sense that intentional states point to antireferents. It remains of course open what they really are, if anything other than precisely the tools available when thinking about others' decision-making.

It is worth to revisit the five assertions to give them a form that is more telling of the position we just took on how intentional states are ascribed, and what good they are in thinking about choices:

1. In order to explain to himself the actions of the observed individual, an observer believes that any individual's actions are interdependent with the intentional states that the observer ascribes to the observed. When the observer uses the concept of intentional state – when he instantiates it and ascribes its instance to the observed individual – he does so to refer via these instances to patterns of behavior that he assumes this observed individual follows.
2. The observer believes that at least some actions of the observed individual follow patterns of behavior that he ascribes – via intentional states – to the observed individual. If the individual behaves in a manner inconsistent to the expectations of the observer, the observer will ascribe to that individual other patterns of behavior, that is, other intentional states.
3. The observer believes that the observed individual has at least some autonomy of choice, so that at least some observed actions of the observed individual are the result of his choosing to follow one pattern of behavior over another. The observer will thus explain the choice of the individual via (1) the intentional states he ascribes to the observed individual, and (2) the reasoning that he assumes this individual performed when making the observed decision.
4. Because the observer explains, rationalizes to himself the choices of the observed individual via the ascription of intentional states to the latter, the decision information ontology of a model of choice should include concepts that capture all kinds of intentional states that he can ascribe. If a model of choice uses that ontology of decision information, is that model necessarily a model of someone else's choice, never of the observer's choice? No: the observer ascribes intentional states to others through language; because he must use that same language – namely, spoken natural language – he can only rationalize, explain

his own choices to others by ascribing intentional states to himself (e.g., he did so and so because he believed this and that). To the extent that the observer has no privileged position with regards to the observed, he has at his disposition the same tools – acts and language – as those available to the observed individual. The decision information ontology that encompasses intentional states is consequently appropriate to rationalize and predict both the choices of the observer and of the observed.

These four may seem quite different from the four assertions initially made above. Roughly speaking, differences come from stating who asserts what. In the initial formulation, the intentional states seemed somehow to be “in” the individual; he “has” or “holds” intentional states, and his actions are interdependent with these, whatever they are. It is as if these initial four assertions were written by a privileged observer, who can see inside the observed individual. It is only from such a position that it could have been claimed – as in the initial fourth assertion – that a decision information ontology should include concepts that capture *all* kinds of intentional states. The revised assertions are much less ambitious: they come from the tenable position that if there is something called “intentional state” then this is something, quite literally *created by the observer* to explain to himself the actions of those he observes. The referent of, say, “belief that *p*” is then some pattern of behavior that the observer ascribes to the observed individual. By ascribing that pattern of behavior to his target, he comes to expect that target to behave in some way, when that target can establish that some conditions are satisfied. What specific behavior that pattern may involve is not something universally established, but depends solely on the pair of (1) the intentional states ascribed to the target, and (2) the conclusions that the observer draws from reflecting on how he would act if he were to ascribe to himself these intentional states he has ascribed to his target. If *A*, the observer believes that a belief in a deity always goes together with praying on Thursdays, and if he finds out that *B* believes in a deity (e.g., *B* had asserted belief in a deity), then *A* may conclude that *B* prays on Thursdays.

The fifth assertion made earlier is rewritten as follows:

5. It is on the basis of communication (both verbal and nonverbal) with the observed individual that the observer chooses the intentional states that he then ascribes to the observed individual.

It was initially said that communication in all of its forms *reveals intentional states*. Presumably, an intentional state can be revealed only if it is somehow “part of” the individual doing the communication and is otherwise inaccessible, hence amenable to revealing through some process, one of communication in particular. That assertion was thus also made from the untenable privileged position, just as the other initial four. In its rewritten form just above, something quite different is asserted, namely, that the observer takes cues on what intentional state to ascribe from the verbal and nonverbal communication, the speech and other actions that the observed individual performs.

### 5.4.2 Decision Information Ontology

Intentional states are ascribed by the observer to the observed individual, the target of advice. The observer does so to explain to himself the current choices and behavior and anticipate the future behavior of the target. The very act of ascription comes out of communication, the observer ascribes intentional states on the basis of the content and mode of communication that he receives from the target.

If the explanation of others' choices does involve the ascription of intentional states, it alone is not enough to explain their choice. The observer first ascribes the intentional states to the observed individual. The observer then needs to assume how the decision-maker processes these intentional states. Since the actual reasoning that the decision-maker performs can only be reflected through communication – just as intentional states are – the observer is effectively forced to speculate how that reasoning may proceed, that is, how the decision-maker organizes these supposed intentional states into a picture of the decision problem he is facing, and how he goes from that picture, that is, what decision rule he applies to select a course of action. The point is that the explanation of another's choice involves speculation about what decision information that observed decision-maker has, how he organizes that information into a concrete decision problem, and what decision rules he applies when seeking solutions to the Decision Problem.

To accept that *the scope of a decision information ontology should cover all intentional states* is to accept that the very language we use reflects our tendency to explain our own and others' choices by ascribing ourselves and others intentional states. Any speculation about others' choices is consequently organized via an engineered ontology, which is reflected in the very way people speak of why they chose in one or another way. This simple observation alone leads to a significant departure from how the standard decision information ontologies are engineered. It is not clear how the primitive terms *alternative*, *outcome*, *payoff function*, and *probabilistic outcome function* in decision information ontologies of perfect and bounded rational choice relate to the terms people seem to use to explain own and others' choices. Perhaps something akin to a probabilistic outcome function and payoff function is indeed what a gambler at a roulette table might use to explain the choice to bet on some numbers instead of others. This might well have been the case if we could ask Joseph Jagers, an engineer who won a significant sum at a casino in Monte Carlo in 1873 by first identifying, among six roulette tables one that had a biased wheel. According to the story, it is with the help of six assistants that he recorded every result (i.e., the number) on six roulette tables for 6 days. He managed to identify from the data one roulette table, the wheel of which was biased giving more often a subset of the 37 possible numbers. Understanding that this gives him a serious advantage, the legend says he ended up winning a sum so large, that he never needed to go back to his previous profession.

The various decision information ontologies have of course been constructed via the usual process: start from some intuitions, find terms that are used to describe those intuitions, see how these terms fit into a decision problem, and see if all of that is useful to either describe choice behavior or prescribe it. As discussed earlier,

in von Neumann and Morgenstern's *The Theory of Games and Economic Behavior*, the initial intuition was that economic transactions resemble interactions between the players of a game. The decision information ontology of game theory consequently arises out of an analogy, and incorporates terms that people commonly use when speaking about games, such as chess, tennis, rugby, and so on.

To say that a decision information ontology arises out of the actual use of language begs the question of why deal with analogies? That is, why go for analogies if people already have a way to speak about their decision-making? By taking an analogy, and using the terminology applicable to the analogy resembles taking a long path to a model of choice, when a shorter one is neither unknown, nor somehow blocked. The answer has to do with predictive power, but that does not seem to be a strong point for choice models based on expected utility theory (Schoemaker 1982). If to speculate about another's choice, and to explain own choice both involve the use of a specific terminology which uses terms pertaining to Intentional states, then it seems reasonable enough to try to use that, so to speak "language of choice" when designing a decision information ontology. That language of choice, i.e., the part of the natural language used commonly to speak about choices undoubtedly developed over a long time, and would not have remained in use if it was dysfunctional, that is, if it made it very hard or impossible to tell and write explanations of others' and own decision behavior. If this is admitted, then terms such as beliefs, desires, intentions, and so on may well have a relevant role to play in an ontology of decision information.

### Scope of the Ontology, and Cues

It was argued in Chap. 3 that the design of an ontology involves a number of ontological choices, the initial ones being made to delimit the scope and depth of the ontology in question. Following the reasoning above – on the role of ascribed intentional states in the explanations of others' choices – the scope of the decision information ontology is determined by intentional states: the ontology should incorporate all concepts needed to cover all intentional states that an observer may ascribe to the observed decision-maker. As a practical consequence of this, the ontology should carry a catchall concept, the extension of which will include *any* particular intentional state. It is not difficult to see that the extension of that concept has exceptionally many members: if Searle is followed, then any instance of the catchall concept is of the form  $m(p)$ , where  $m$  is the psychological mode verb and  $p$  some content. It is evident from natural language both that (1) there are many verbs for different psychological modes and that (2) the object these verbs are associated with can be a wide variety of sentences.

It is not perfectly clear what are the essential properties of the content of an intentional state, or equivalently, there is no unique definition for the content of intentional states. The list of the essential properties that some sentence  $p$  must satisfy to be admitted in  $m(p)$  is not settled. The content of an intentional state is often called a *proposition*, which is a technical term (McGrath 2008; Crane 2001) having



different definitions: a proposition may be a sentence having a truth value, i.e., a bearer of truth value, or any object of a verb used to tell of a psychological mode. It is a debatable issue whether all contents of intentional states must have a truth value, whether it is possible to say for any proposition that it is either true or false. Loosely speaking, the idea on truth values is this: the truth value of a proposition depends on whether the conditions expressed in the proposition (i.e., the properties and relations stated for some objects) “correspond to reality”: e.g., the proposition “the door in front of me is open” is true if that door that I can see in front of me indeed is open. If any content of an intentional state must have a truth value, then this value seems unclear if, e.g., a sign in that proposition has an antireferent. There are, however, people who do actually believe that, e.g., the proposition “Afghanistan became a democracy in 2009” is true, where “democracy” targets an antireferent, and certainly so when combined with the term “Afghanistan.” The requirement that a proposition must have a truth value is one that many contents of intentional states cannot satisfy. This lack of clarity about the properties of the content of intentional states leads to a further ontological commitment, namely, that any content of an intentional state gets called a proposition here, all the while it need not have a truth value: the term *proposition* will be used interchangeably with the phrase *content of intentional state*, and will be any sentence that can be the object of intentional states. This is clearly not a good definition, and *proposition* consequently remains a primitive term here.

The extension of the catchall concept in this decision information ontology will certainly be large: not only are there many verbs for psychological modes, but it seems that anything that can be the object of any of these verbs is a proposition. This last remark is not acceptable, however: when deciding what flight ticket to buy, one’s belief that his car is red seems irrelevant to his decision. A decision situation, or a decision setting is simply another name for a context of reference, the elements of which are those objects that satisfy the conditions discussed in Chap. 3. It appears then appropriate to say that an intentional state should be relevant to the decision problem, and it will be so if its proposition refers to some object in the context of reference, i.e., the decision situation. It follows that *not any* intentional state will be relevant for the decision-maker. Let us call cue this catchall concept<sup>5</sup> and consider the definition:

**Definition 5.1. Cue:** Any instance  $c$  of the concept cue for the observer  $\mathbf{O}$  of the decision-maker  $\mathbf{D}$  must satisfy the following identity criteria:

1.  $c$  is a particular intentional state  $m(p)$  that  $\mathbf{O}$  ascribes to  $\mathbf{D}$ .
2.  $\mathbf{O}$  believes that  $\mathbf{D}$  can form reference relations between  $p$  and objects in  $\mathbf{D}$ ’s context of reference.

---

<sup>5</sup> As a notational convention, we will write **cue** in a sans-serif typeface to refer to the concept, and cue in a serif typeface to refer to any instance of cue.



The first condition above is curious: it is given not by considering the decision-maker alone, but from the perspective of an observer. It begs the question of why something like the following is not offered instead:

Any instance  $c$  of the concept **cue** for the decision-maker **D** must satisfy the following identity criteria: (1)  $i$  is a particular intentional state  $m(p)$  that **D** ascribes to himself; (2) **D** can form reference relations between at least some element of  $p$  and objects in **D**'s context of reference.

This second proposal gives the idea that when an individual chooses, he attributes himself desires, beliefs, intentions, and so on, and perhaps then reflects thereon before choosing. That is without a doubt a nice picture, showing the decision-maker who seems to step out his own skin to ascribe himself and look upon his own intentional states, consider them from this other perspective, and then choose. While it is certainly possible to reflect on one's beliefs, intentions, and so on – i.e., have beliefs about beliefs, beliefs about desires, etc. – we should recall that intentional states are what we observe from the use of language, and not from introspection. It is consequently interesting to avoid matters of introspection and hence safer to distinguish two roles, the observer who ascribes the intentional states and the observed decision-maker. If and when the individual does ascribe to himself intentional states and reflects thereon, he is occupying both of these roles.

The definition of *cue* says that any instance thereof is simply an ascribed intentional state, the content of which refers to objects in the decision-maker's context of reference. Just as when a definition of advice was being built, it is relevant to ask how the *cue* concept fits within a foundational ontology. Is it a specialization of a perdurant, endurant, of a quality, or of an abstract? Since cues are intentional states satisfying the second condition in the definition above, the question is rather where intentional states fit with regards to the basic classes of particulars in an engineered foundational ontology. The term itself, intentional *state* indicates that particular intentional states are perdurants, but instead of being processes, they are states (Crane 2001); it seems less intuitive to say that "a belief that  $p$ " is a process, than a state – coming to believe that  $p$  seems a process, the result, namely, "belief that  $p$ " is a state. "Believing that  $p$ " is then a process, through which the state "believe that  $p$ " is maintained.

Why use the term *cue* as the name of the catchall concept in this decision information ontology? *The Oxford English Dictionary* says that in theater, *cue* was used since the sixteenth century to refer to the "the concluding word or words of a speech in a play, serving as a signal or direction to another actor to enter, or begin his speech," and that in music, a cue is "a direction to enable a singer or player to come in at the right time after a long rest"; in its more recent use *cue* refers to "a stimulus or signal to perception, articulation, or other physiological response." What is then the link between cues and intentional states? The observer will ascribe intentional states to the observed decision-maker to explain his behavior. An intentional state is thus something that the observer will take as a condition for some behavior – i.e., "he behaved in such and such way, because he believed this and desired that." In this respect, the ascribed intentional state resembles a cue, a something that can be a condition for, and can result in an action, a choice. If an individual sees a closed

door, thus presumably believes that the door is closed, and desires to go through that door, the observer may expect him to act in some predictable way to open that door so that he can pass through it. From the perspective of the observer, the beliefs and desires he ascribes to the decision-maker amount to cues for that decision-maker, for it is only as cues that they are relevant to explain anything. This of course does not mean that people are conditioned as Pavlov's dogs, but only that from the perspective of the observer, they do appear to act because of some complex of intentional states, or in other words, that some complex combination of intentional states does precede and condition what actions are taken.

That any cue can be written as  $m(p)$  points to two ways to specialize this catchall concept. One option is to specialize according to the properties of the content of the Intentional state, namely the properties of  $p$ ; the other option is to specialize according to the properties of  $m$ . The two options are not incompatible. Following the distinctions made above between beliefs, desires, intentions, and evaluations as kinds of psychological mode, the second option will be taken now to specialize the cue concept.

### Beliefs and Desires

The phrase *belief that  $p$*  is usually considered as referring to the intentional state that an individual has when he regards  $p$  as being true. Care should be taken in what is read from "true": the cautious position is that an individual can have false beliefs, but which he may regard as true at some point in time, when facing some choice. It is, in other words, too much to ask that one only believes statements that have truth as their essential property. Recall what was said of truth and rational justification in Chap. 3: given some beliefs that the individual may hold, some of these beliefs may be rationally justified for him to believe as true, but may be false. Some of the beliefs he holds may have been proved false, so to speak, once and for all: e.g., he believes that planet Earth has two moons. Other statements may not have truth or falsity as their essential property, but can be still be beliefs of an individual. Needless to say that there are different positions and conceptions on what *belief that  $p$*  exactly refers to (Schwitzgebel 2008).

According to one view (Fodor 1975), to say that "**D** believes that  $p$ " refers to **D** having stored  $p$ , not unlike a computer, and being capable to use  $p$  to perform some restricted set of operations: it is then precisely the set of operations that can be effected on  $p$  that determines if  $p$  is a belief or something else. Along this line of thinking, some operations can only be performed for beliefs, others only on desires, and so on. Hence, **D** believes  $p$  if he has, so to speak, recorded  $p$  among those other information to which he can apply belief-specific operations.

Another view is, roughly speaking to say, that the essential properties of an intentional state are its causal relations to sensory stimulations, behavior, and other intentional states (Putnam 1975). An individual looking at a closed door in front of his may believe that there is a door in front of his and that the door is closed; if he desires to go through the door, that belief will play some role in the formation of his

intention to open the door before going through it. The particular belief “there is a closed door in front of me” of this individual is thus defined via the sensory inputs – his perception of the closed door – and the effects that this belief may have on his to entertain other beliefs (e.g., concerning whether the door can be opened, whether it is locked), other intentional states, such as desires (e.g., whether he desires to go through the door) and intentions (e.g., whether he intends to go through the door), and through his intentions, on his behavior. The rough idea is thus to see particular beliefs in terms of the role they play in relating sensory inputs and behavior.

Yet another perspective is that of an observer of **D**: **D** believes *p* according to that observer if it is *prima facie* evidence to the observed that **D** behaves as if **D** believes *p* (Dennett 1978). This is very much in line with the ideas discussed earlier, that intentional states are ascribed to explain behavior. Belief is then a psychological mode, the role of which is to condition behavior in some ways different from how other intentional states condition behavior. The observer who ascribes a particular belief to another individual will expect that individual to act in some way, a belief-specific way as opposed to a desire-specific way, an intention-specific way, and so on.

It might seem that a particular belief conditions behavior in such a way that beliefs should not be violated. If one ascribes to another a belief, this story goes, then that other will not choose to act against that belief. It is not difficult to see the problems with this position. First, the individual who has been ascribed the belief may be unaware of that belief in the decision situation; there are beliefs that may have been held previously, but are simply not recalled when making some choice. Second, an ascribed belief is just that: a belief that has been ascribed by someone else to an individual, and hence need not be “held” by that individual at all. Third, perhaps there are other beliefs that can be ascribed to that individual, and that may conflict with the one ascribed initially – which one of these will he violate? We will revisit this last question later on. For now, we introduce the concept of assumption in our decision information ontology, and do so in a way that is neutral on what an observer will conclude from an assumption that he has ascribed to the observed individual.

**Definition 5.2. Assumption:** Any instance of the concept assumption for the observer **O** of the decision-maker **D** must satisfy the following identity criteria:

1. *a* is a particular belief-that-*p* that **O** ascribes to **D**.
2. **O** believes that **D** can form reference relations between *p* and objects in **D**’s context of reference.

The definition simply says that an instance of assumption is an instance of cue, in which the psychological mode is that of belief. Assumption is thus a specialization of cue; the extension of the former is a subset of the extension of the latter concept.

In contrast to beliefs, the story of desires is that the individual desiring something is inclined to act in ways that may satisfy the desires. This is certainly a simplistic view, but it is good for now, as we are not concerned for the moment with the dynamics of intentional states (i.e., what one will or should do given some set of

ascribed intentional states). If the individual is ascribed a belief-that- $p$ , then he is assumed to think that  $p$  describes the state of affairs as it is. In contrast, if we ascribe him a desire-that- $p$ , then it seems according to the simplistic view that he does not see that  $p$  describes the state of affairs, but that he deems desirable the states of affairs described by  $p$ . In any case, goals will be here kinds of cues, in which the psychological mode is that of desire.

**Definition 5.3. Goal:** Any instance of the concept goal for the observer **O** of the decision-maker **D** must satisfy the following identity criteria:

1.  $g$  is a particular desire-that- $p$  that **O** ascribes to **D**.
2. **O** believes that **D** can form reference relations between  $p$  and objects in **D**'s context of reference.

An important consequence of our taking the observer's view is that the definition of goal says nothing on if a goal being judged appropriate or not, according to, say, an ethical or moral principle. The latter concern the evaluation of a future state of affairs, perhaps in relation to another potential future state of affairs. The goal concerns a potential state of affairs that the observer believes the target desires, or evaluates as more desirable than another potential state of affairs. Whether what the observed individual is ascribed as a desire is desirable to the observer is in the eye of the observer. It is critical to see this separation between potential future states of affairs, ascribed belief in the realization of these states of affairs, and the evaluation of the potential future states of affairs. As soon as we see that separation, we get to an important point about how ascriptions of desires operate: to ascribe a desire, the observer needs to ascribe to the other (1) belief in the realization of potential future states of affairs, and (2) the evaluation that the observer may have, in terms of desirability of these realizable states of affairs. To ascribe desires is to ascribe quite a lot then, and it is not strange that such ascriptions go very wrong when done by individuals unaware of contexts of those whom they ascribe desires. It is then not very strange that a soldier sent from afar to an occupied territory feels betrayed when being shot at, having been sent to help precisely those who are shooting. Frustration might be avoided if it was clear to this soldier that he is ascribing desires on grounds of what he himself evaluates as desirable.<sup>6</sup> This conveniently leads us to evaluations.

---

<sup>6</sup> Reports of the feeling of betrayal are explicit and recurrent as far as the veterans of the 2003 and ongoing Iraq war: e.g., "I guess while I was there, the general attitude was, A dead Iraqi is just another dead Iraqi," said Spc. Jeff Englehart, 26, of Grand Junction, Colorado. Specialist Englehart served with the Third Brigade, First Infantry Division, in Baquba, about 35 miles northeast of Baghdad, for a year beginning in February 2004. "You know, so what?... The soldiers honestly thought we were trying to help the people and they were mad because it was almost like a betrayal. Like here we are trying to help you, here I am, you know, thousands of miles away from home and my family, and I have to be here for a year and work every day on these missions. Well, we're trying to help you and you just turn around and try to kill us." (Hedges and Al-Arian July 30, 2007).

## Evaluations

While beliefs and desires are to be found in, respectively, assertive and directive speech acts, the expressive speech acts convey the speaker's attitude, emotion, or feeling. We will call *evaluation* any intentional state that may be conveyed by expressive speech acts; this choice of this term arises from the conceptualization of attitudes, and their relationship to emotions, feelings, and moods in psychology. In order to understand the evaluation concept here, it is necessary to make a detour and consider the notions of attitude and emotion, and the role they may play decision-making.

The notions of attitude, emotion, feeling, and mood are strongly related. Attitude has been defined as "a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor" (Eagly and Chaiken 1996). An attitude thus gives an evaluation in terms of degree of favor or disfavor. Such degrees vary in sign (positive or negative) and in intensity, whereby the intensity of the valuation is relative: considering an object of attitude on its own involves implicit comparison to a set of objects perceived by the evaluator to be of the same kind. "[O]bjects of attitudes [as the term is used psychology] include anything that people can like or dislike, wish to protect or to harm, to acquire or to reject" (Kahneman et al. 1999). Attitudes can arise from any combination of affective, cognitive, and behavioral input (Bizer et al. 2003). Affective basis of attitudes comes from emotions, feelings (i.e., conscious subjective experience of emotion), and moods (i.e., long lasting, less intense, less-specific affective states) associated with the evaluated object through past or current experience. The experience of the object's qualities, i.e., the object's properties give the cognitive basis of attitude. The behavioral basis of attitude is found in the individual's interpretation of past behavior and the intentions to commit future acts. It is further relevant to note that separate attitudes on two objects do not necessarily predict the outcome of a choice or direct comparison of these objects (Hsee 1996; Kahneman et al. 1999). Attitudes are richer and less well behaving in mathematical terms than those of utility. It has been suggested (Kahneman et al. 1999) that the consistency and short-term stability of preferences (Sen 1973) are missing in attitudes. Moreover, attitudes violate what is called "extensionality," because the same object may evoke different evaluations depending on the context in which it is being evaluated (Kahneman et al. 1999).

It is uncontroversial to say that emotions influence decision-making, in the sense that they influence one's evaluations of the courses of action he has considered. Question is: how do emotions influence the ranking of alternatives in a decision situation? To say that they give evaluations is not precise enough. Economists have relatively recently started to study the role of emotions in decision-making. According to Jon Elster's review of the "emotion" concept in economics research, "[by] far the most common way of modeling the interaction between emotions and interests [i.e., values that the decision-maker sees in alternative choices] is to view the former as psychic costs or benefits that enter into the utility function on par with satisfaction derived from material rewards" (Elster 1998, p. 64). If we take emotions to shape the reward parameters in a decision situation, we can say that emotions are a

source of economic preferences. Elster argues, however, that emotions “also affect the ability to make rational choices within those parameters [i.e., the parameters that describe a decision situation]” (Elster 1998, p. 73). Emotions can thus lead to a bias toward particular alternatives, which finds support in neuroscience research (Dunn et al. 2006; Camerer et al. 2005). António Damásio, a neurologist, argues that the mechanism of this bias is as follows (Damasio et al. 1991; Damasio 1994):

[A generic situation is described, in which a choice is to be made.] [...] imagine that *before* you apply any kind of cost/benefit analysis to the premises, and before you reason toward the solution of the problem, something quite important happens: When the bad outcome connected with a given response option comes into mind, however fleetingly, you experience an unpleasant gut feeling. Because the feeling is about the body, I gave the phenomenon the technical term *somatic* state [...] and because it ‘marks’ and image, I called it a *marker* [...] What does the *somatic marker* achieve? It forces attention on the negative outcome to which a given action may lead, and functions as an automated alarm signal [...] The signal may lead you to reject, *immediately*, the negative course of action and thus makes you choose among alternatives. The automated signal protects you against future losses, without further ado, and then allows you *to choose among fewer alternatives*. There is still room for using a cost/benefit analysis and proper deductive competence, but only *after* the automated step greatly reduces the number of options [...] This general account applies to the choice of actions whose immediate consequences are negative, but which generate positive future outcomes. (Damasio 1994, pp. 170–175; Damasio’s emphasis)

According to Damásio’s somatic marker hypothesis, choice behavior combines cost/benefit analysis with marker signals that indicate how rewarding or punishing an alternative course of action is likely to be (Dunn et al. 2006).

The perspective here is one of the observer, who ascribes – in addition to beliefs and desires – emotions, moods, attitudes, and feelings to the other. Compassion, for instance, requires that there is an ascription of specific emotions to the other, e.g., pain or suffering. To answer what role ascriptions of emotion and such have for the observer, it is important to see two consequences of the discussion of emotions, attitudes, feelings, and moods. One is that there are few if any restrictions on what an emotion, feeling, mood, and psychological attitude in general may have as its object, of what it may be about. If the observer ascribes happiness to the other, that alone is not terribly interesting in predicting the other’s behavior. What is of significance is how the observer expects this emotion to influence the behavior of the other; if the latter appears frightened, and perhaps this will lead his to attempt to escape his current situation; if he seems happy, perhaps this will lead his to act to preserve that situation; and so on. This is where the second conclusion comes in, namely that emotions and such seem to act as a filter for information considered in a decision situation: somatic markers act as signals that eliminate some courses of action even before they are being thoroughly considered.

From the perspective of the observer, the evaluations that he ascribes to the other act as a way to organize and perhaps reject other cues that he may have previously ascribed. If the other seems thrilled when being in London as opposed to being in Paris, the observer may conclude that the other prefers being in London to being in Paris. The observer sees the other considering flights, one for London another for Paris, and seeming more interested in the London one, so that he seems to the

observer to desire “more” the realization of the goal of being in London than of the goal of being in Paris. An evaluation in this case compares two goals. In other cases, comparison need not be involved: if the other seems unhappy for being currently in Paris, the observer may conclude that he will desire to leave Paris. To reach that conclusion, the observer would ascribe to the other that (1) an assumption, that he believes he is in Paris, (2) an evaluation of that assumption, that he is unhappy with being in Paris, and (3) a goal, that he desires to leave Paris. If the observer ascribes also the evaluation that the observed individual dislikes Rome, then the conclusion would not only be that the other will have a goal to leave Paris, but that he will have a goal to leave Paris and not go to Rome.

The aim with the evaluation concept is that its instances can be any intentional state that may be conveyed by an expressive speech act, and thus any intentional state that pertains to emotion, attitude, feeling, or mood. If anything can be the object of evaluations, and if they evaluate in terms of desirability these objects, then an evaluation may have any cue as its object. If an evaluation compares two or more cues, that evaluation may thus capture a temporary preference order. Because psychological attitudes violate extensionality [among other typical characteristics of standard economic preferences (Kahneman et al. 1999)], the temporary preference order need not be stable, transitive, or necessarily obey some other property.<sup>7</sup> As it is impractical to enumerate all verbs for psychological mode, and which may be used in an expressive speech act to convey emotion, mood, feeling, or attitude, the definition of evaluation will not enumerate intentional states. It instead says that an instance of evaluation assesses other cues in terms of desirability.

**Definition 5.4. Evaluation:** Any instance  $x$  of the concept evaluation for the observer **O** of the decision-maker **D** must satisfy the following identity criteria:

1.  $x$  is a particular intentional state that **O** ascribes to **D**.
2. the content  $p$  of  $x$  refers to one or more other cues.
3.  $x$  evaluates in terms of desirability the cues referred to in  $p$ .
4. **O** believes that **D** can form reference relations between  $p$  and objects in **D**'s context of reference.

## Commitments

With assumption and goal, the decision ontology covers two main classes of ascribed intentional states, namely beliefs and desires. We say classes, because it is plausible as Searle argued that some intentional states can be reduced to a combination of desires and beliefs (Searle 1983). With the evaluation concept, the idea was to cover

---

<sup>7</sup> For such an evaluation to be unstable means simply that it need not be kept over time or across contexts: if one prefers today vanilla ice cream to that of chocolate, there is little to oblige him to continue preferring the former to the latter in the future. To say that the evaluation need not be transitive is that if one prefers  $x$  to  $y$  and prefers  $y$  to  $z$ , that he necessarily also prefers  $x$  to  $z$ .



intentional states that evaluate in terms of desirability some cues, and which are communicated via expressive speech acts.

What seems to be irreducible to only beliefs, desires, and evaluations is the notion of commitment to act, the idea that one makes a choice to act in some way.

To take the perspective of the observer who ascribes intentions to the observed individual to predict his behavior is to be concerned with intentions directed at future action. Michael Bratman, a philosopher, suggested that such intentions are building blocks of plans of action (Bratman 1987). He argues that intentions are not easily changed, they remain stable and condition the formation of future intentions. They thus help organize behavior over time. While desires may orient behavior, intentions involve commitment to act, the relation between intentions and actions is thus in a sense tighter than that between desires and actions. *While desires seem to orient actions, it still remains to the individual to choose which desires to satisfy and how; if he holds a future intention, then he must already have made the choice of which desires to satisfy, and how to do so.* Commitment is an essential property of intention, according to Bratman, in which it imposes some regularities on behavior, namely (1) to revise intention only in case relevant new information comes to light, (2) to engage in reasoning about the means necessary to achieve the intended and desired ends (i.e., to see what to do), and (3) not to consider actions that counter the realization of the intention.

Independently on how intention may regulate behavior, to ask what intentions are requires an understanding of how intention differs from plans of action. If a plan is only some – more or less precisely specified – sequence of actions, e.g., as cooking instructions are, then intentions are certainly not plans. In that sense, plans as recipes-for-action are merely guidelines on what to do: it remains open who may want to do so, whether they believe that a plan can be performed successfully in a given context, and so on. There is then a difference between a plan-as-a-recipe-for-action and the term *plan* when we use it to say that some individual “has a plan” – in the former case, the recipe-for-action only says what to do regardless of whom, when, where would be doing it. In the latter case, when someone is said to have a plan, it is assumed that *there is commitment to act in some chosen way*. This latter view (Pollack 1990) accords with Bratman’s and is that plans are not combinations of act descriptions, but combinations of intentional states. In contrast to being merely instructions on what actions to take, this other view of plans incorporates the individual acting according to plan, or rather, speaks of an individual as having a plan: *to have a plan is then a relation between a recipe-for-action and the individual who has adopted that recipe-for-action. To adopt a plan is to commit to it, and hence, to have a plan is to have future intentions.* To have a plan also requires beliefs, in the sense that one needs to believe that performing the actions in a recipe-for-action will give some anticipated effects. Moreover, the individual will need to believe that each of these actions does indeed play a role in the plan, that they all need to be performed in some either known or figured-out-along-the-way sequence to obtain the end effect.



Taking then the perspective of the observer who ascribes intentional states, to ascribe a plan – to say that *the observed individual has a plan* – is to ascribe the following:

1. Assumptions, to reflect beliefs-in-actions that the observer ascribes to the other. Beliefs-in-actions themselves come in three kinds: beliefs (1) in what conditions must be met to perform an action, (2) that these conditions can be met, and (3) that the performance of the action will produce desired effects.
2. Goals to account for the desires that the observer believes are directing the behavior of the observed individual.
3. Commitment to perform some of the actions, for which the observer had ascribed to the observed individual some beliefs-in-actions. When ascribing commitment, the observer believes that the observed individual has chosen to follow some recipe-for-action, or if the recipe is not sharply defined, then design and obey it along the way.

If one needs to ascribe the above to ascribe a plan, then *plan* is hardly a primitive term, but is definable in terms of assumptions, goals, and commitments. Hence the introduction of the commitment concept, as we already have assumptions and goals.

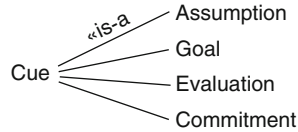
**Definition 5.5. Commitment:** Any instance of the concept commitment for the observer **O** of the decision-maker **D** must satisfy the following identity criteria:

1. It is a choice that **O** ascribes to **D**, by which **O** chose some assumptions and goals.
2. **O** believes that **D** can form reference relations between *p* and objects in **D**'s context of reference.

To say that an individual has a plan is to ascribe beliefs-in-actions, desires, and commitments. Ascribing a plan is a shortcut. To ascribe a commitment amounts to ascribe the other a choice of some assumptions and goals over others. Hence, to ascribe a plan is to ascribe a commitment to some set of beliefs-in-actions and desires, and not in others.

### 5.4.3 *Synthesis of the Decision Information Ontology*

The decision information ontology includes five concepts (Fig. 5.2): cue and four concepts that are its specializations – assumption, goal, evaluation, and commitment. All concepts are defined from the perspective of the observer of a decision-maker. Assumptions and goals are, respectively, the beliefs and desires that the observer ascribes to the decision-maker. Evaluations are intentional states pertaining to emotions, feelings, moods, and attitudes that the observer ascribes to the decision-maker. Finally, commitments are intentions that the observer sees in the decision-maker. There are no probabilities and no utilities; their less prominent role will become clearer in the next chapter.



**Fig. 5.2** Taxonomy of the decision information ontology, i.e., the specialization of the **cue** concept. Every line in the figure refers to the specialization (is-a) relation

What about the actions that the observed decision-maker may be capable of – how do they fit in this picture? From the observer’s standpoint, if the other acts according to plan, then he is performing and will perform actions. In order to anticipate which actions will be performed, i.e., what plan will be followed by the other, the observer will ascribe something that could be called *capabilities*. Capabilities are in a sense potential actions. If the observer ascribes a capability to the other, then he believes that the other could perform some specific action if he chose to do so (i.e., if he committed to do so). What is being ascribed then, when a capability is ascribed, or in other words, when we assume that someone else can perform some action? As a future action is obviously an action that has not yet been performed, to ascribe a capability of executing that action in the future is to ascribe beliefs-in-that-action. If the observer **O** ascribes a capability (to execute the action) *a* to the decision-maker **D**, then **O** ascribes to **D** beliefs of four kinds, as indicated in the following definition:

**Definition 5.6. Capability:** Any instance *x* of the concept capability for the observer **O** of the decision-maker **D** must satisfy the following identity criteria:

1. *x* is a set of particular beliefs that **O** ascribes to **D**.
2. Any particular belief in *x* is either:
  - (a) That some specific conditions must be met to perform an action
  - (b) That these conditions can be met (by **D** alone, with others, etc.)
  - (c) That performing an action produces some specific effects
3. **O** believes that **D** can form reference relations between *p* and objects in **D**’s context of reference.

When the observer ascribes a capability, he ascribes beliefs to the decision-maker, whereby these beliefs are about some action. As the action is potential, it *may* be (not *is*) performed, to ascribe capability is to ascribe a set of beliefs about that action. These ascribed beliefs fall into three categories: (a) beliefs that the action, say *a*, can only be performed when some specific conditions hold; (b) beliefs that these necessary conditions can be met, and this by the decision-maker alone, in coordination with others, and so on; and (c) beliefs that performing *a* results in some specific effects. In simpler terms, when a capability of performing action *a* are being ascribed, beliefs needed before committing to that action *a* are in effect being ascribed. As an aside, note that an instance of capability is thus a set of instances of

the assumption concept, whereby only some assumptions are admitted. To anticipate actions, the observer ascribes capabilities, and such ascriptions amount to those of beliefs. Capability thus remains outside the decision information ontology, but will be useful to highlight that some assumptions are particularly relevant to the prediction of future actions.

Capability is not a primitive concept here, not a specialization of cue, but a construct built from the primitive concepts in the Decision Information ontology. Other constructs can be defined in the same manner. We said, in discussing commitment that to ascribe a plan to another, the observer will ascribe beliefs-in-actions (i.e., capabilities), desires, and commitments. A plan then, as in “this person has a plan” is a combination of beliefs-about-actions, desires, and intentions. To ascribe an instance of plan is to ascribe a set of instances of the capability, goal, and commitment concepts. If we take an instance of plan, and deprive it of the instances of commitment and goal, we get a recipe-for-action, i.e., a set of beliefs-about-actions, or in yet other words, a set of capabilities.

This synthesis of the Decision Information ontology closes this section. We started it off with the observation that there are three parts to a model of choice. One is a Decision Information ontology, the role of which is to give concepts serving to categorize information relevant to a decision-maker. Not only does it posit that all that is relevant is in the ontology, or can be built from the combination of the concepts in the ontology, but also that all that is outside of the ontology, or cannot be made from its ingredients is effectively irrelevant.

The other two parts are the decision problem and the decision rules. In a given decision situation, the decision problem should tell the decision-maker how to relate the instances of the concepts from the Decision Information ontology in order to have an organized, as opposed to chaotic picture of the decision situation. While the Decision Information ontology serves as a filter in looking at a decision situation, the decision problem organizes what the filter lets in. The decision rules are the properties that a solution to the decision problem should have.

## 5.5 Taxonomy of Advice

The five concepts – cue, assumption, goal, evaluation, and commitment – give a way to speak about decision situations, forming the foundation of the language that an observer may use to rationalize, explain the behavior of the decision-maker who is being observed.

### 5.5.1 *Whose Explanations?*

The five concepts arise out of taking the perspective of the observer of a decision-maker, and neither of some detached speculative inside (i.e., introspective) position,

nor some above and below (i.e., omniscient) place. It is thereby *not* a position of authority, from which one could claim to understand some essence of the decision-maker. Perhaps, it may/will be in the future shown to be truly relevant to say that “person  $x$  has a belief that  $p$ ” in the sense that saying this somehow captures something proper to person  $x$ , some quality that is inherent to that person and can be recognized in the same way regardless of the observer. If that was to be shown in the future, then such a statement would be alike one for planet Earth, that “Earth exerts gravitational force.” The point in this parallel is not that belief should be somehow a stable quality of a person, as the exertion of gravitational force is for Earth, but that it is really not clear at all what quality of a person  $x$  one is referring to when one says that “person  $x$  has a belief that  $p$ .” It is, however, clear what we are referring to when we say that “Earth exerts gravitational force”: the statement is a conclusion of a procedure that has been repeated countless times and had given the same outcome; the procedure itself is understandable in the same way by many, who may choose themselves to perform it to verify whether and how strong gravitational force is being exerted. There is no such procedure that different people at different times can apply to agree that “ $x$  has a belief that  $p$ ,” or if  $x$  “changed his belief,” then to verify whether “ $x$  had at time  $t$  a belief that  $p$ .” The closest procedure seems to be one that uses language: namely, ask the person if he believes that  $p$ . Lying of course makes this an unreliable method. Questions augmented with torture may seem to the barbarian a way out, but this is no good, being ethically unacceptable. What solid, stable remains then of all that talk of intentional states?

Once the position of an introspective or omniscient observer is rejected in favor of the much more humble one, namely of a partner in communication, what remains of intentional states is that they are simply something in natural language that people use to explain others’ decisions. The decision information ontology reflects this facility used to organize the thinking about others’ choices. The language that uses terms for ascribed intentional states is a tool in the sense that these terms categorize one’s experience and intuition about the other. Presumably then, people distinguish belief from desire because the experience and intuition about the former is different than that for the latter. This tool – the part of natural language for talking about ascribed intentional states – seems to have changed little since Sophocles, as Paul Churchland, a philosopher, argues (Churchland 1981). The little change should, according to him, certainly not be read as an argument that common language of intentional states is the best of potential tools. It would be heroic to anticipate that we cannot find a better one for the construction of explanations of others’ behavior. When perhaps neuroscience replaces the ascribed intentional states with something else, the decision information ontology here should be replaced as well. Before that time, we are stuck with the language of ascribed intentional states, however, backward this may be.

We said at some earlier point that a decision model and hence a decision information ontology it uses arise out of a model of man. Once it is accepted that a model of man requires looking into “the mind of man” and this being hard to access, we need some more accessible candidate. Hence the interest in a model of how people interact via communication. Language use is accessible to experience, while whatever

may be “in the mind” (e.g., an individual’s “real” beliefs, desires, etc., or some other structures) is much less so. If it were better accessible, an authoritarian position could be justified: we would know (better than the observed person himself) how the mind works, and from such a model construct the needed ontology. As this is not a plausible option, we are interested in *the talk of* intentional states, the talk through which one rationalizes the acts of another. Such talk ascribes intentional states, and the decision information ontology merely reflects a rough classification of the variety of intentional states that may be ascribed when the explanations for choices of others are being constructed. The point is that the individual we are observing – who through speech and other acts communicates with us – thus does not *have* intentional states. When the observer says that they have, the observer is simply using language to organize his own explanations of the past, present, or future (predicted) behavior of that other individual. This talk of intentional states is thus *not interesting because it says that people have beliefs, desires, and so on: rather, it is interesting only because those ascriptions indicate that there are different kinds of explanations that people form when they try to figure out why others have been doing something, why they are currently doing whatever it is that occupies them, and why they may do something in the future*. The observer will anticipate different future behavior if he ascribes a desire-for-*p* then if he ascribes the belief-that-*p* to the other. The very point of the decision information ontology is consequently *not* to put beliefs, desires, intentions, and so on into the mind of the decision-maker, to see the decision-maker as being a complex web of intentional states: it is instead to say that there are different elements to – used in – explanations of other’s choice behavior, and that the distinctions between desires, intentions, and other intentional states *in the language* reflect the differences between the elements used in an explanation. Ascribed desires, beliefs, and so on are hence not “in the mind of” the observed decision-maker: instead, our decision information ontology says only that they are in the explanation of the observer.

### 5.5.2 *Specialization of the Concept of Advice*

To obtain concepts for an ontology of advice from those of the decision information ontology requires simply that we recall what led in the first place to the construction of the latter ontology. The story was as follows. Advice, we argued, should be categorized according to what it targets. Potential targets are the various information that may be relevant to the decision-maker, whom we are advising. This is why our earlier discussion of the kinds of advice led us to review various models of choice. Each model of choice incorporates a decision information ontology, the purpose of which is to classify the information deemed relevant to the decision-maker. If the decision information ontology includes the concept of *alternative*, then a corresponding ontology of advice should include *advice about alternatives*, which would be advice dispensed to influence the decision-maker’s understanding of which alternatives are available in his decision situation. The trouble was that it makes little

sense to dispense advice on probability and utility values. This made any model of choice that incorporates these two notions irrelevant to the engineering of an ontology of advice. It only follows in such a situation that we engineer first a decision information ontology, and from there on design an ontology of advice. Hence our present preoccupation with the ontology of advice, given that we now have a decision information ontology which is intolerant to substitutes.

The individual whom we called the observer **O** earlier now becomes the advisor, and we will denote him **A**. **D** remains the observed individual, the decision-maker who is being advised by **A**. The idea is to specialize advice according to its target: e.g., if **O** ascribes a belief to **D**, then there is a kind of advice that will target that belief, to change it. Advice that will target ascribed beliefs can thus be distinguished from one that will target the desires that **O** ascribes to **D**, just as advice targeting ascribed emotions, moods, and so on, is distinct from the advice that targets ascribed intentions. We will now repeat the definition of advice and introduce some additional notation to facilitate further discussions.

**Definition 5.7. Advice:** Any tuple  $(A, x, D)$  is an instance of advice if and only if it satisfies the following identity criteria:

1.  $x$  is some potentially complex speech act performed by **A**.
2.  $x$  is experienced by **D**.
3. **D** can distinguish the dicta from modi in  $x$ , and from the modi establish which of the dicta he could adopt as beliefs, desires, intentions, or evaluations.
4. **D** can form reference relations between at least some of the dicta in  $x$  and objects in his context of reference.

Advice that targets a cue is given either to support the cue or to attack it, hence the following definition of the advice-about-a-cue concept:

**Definition 5.8. Advice about a cue:**  $(A, x, c, D)$  is an instance of advice-about-a-cue if and only if:

1.  $(A, x, D)$  is an instance of advice.
2.  $c$  is an instance of cue that **A** holds about **D**.
3. According to **A**,  $x$  either supports or attacks  $c$ .

The idea in the above definition is this: the advisor communicates advice  $x$  to the decision-maker to either support the cue  $c$  or attack it. Suppose for example that **A** believes that (1) **D** wants to open the door, (2) **D** does not have the key to open that door, and (3) the door is locked. Let  $c$  be the desire to open the door that **A** ascribes to **D**. To support  $c$ , **A** could tell **D** where the key is, or how to open the door without the key. Such advice is indeed supportive, or favorable as **A** is recommending a course of action that will not deny  $c$ . If **A** is instead willing to attack  $c$ , he could tell **D** that he cannot go through that door, or that **D** should go through another door.

It is useful to recall here that **A** works by speculation and simulation. **A** gives advice to influence the intentional states that **A** ascribes to **D** – whether these intentional states truthfully describe **D** is a different problem. **A**'s ascription of intentional

states to **D** is the first step in advice giving that involves speculation. The second step is for **A** to anticipate what **D** may commit to doing if the ascribed intentional states, i.e., cues that **A** has about **D** were indeed a proper description of **D**. That second step, the anticipation is speculation as well, but it involves simulation. The third step sees **A** designing advice that he will give to **D**, to influence **D**'s behavior, so that it either follows or departs from the predictions that **A** obtained via simulation.

It is straightforward to specialize the advice-about-a-cue according to the taxonomy of cues. The result are concepts of advice-about-an-assumption, advice-about-a-goal, advice-about-an-evaluation, and advice-about-a-commitment.

**Definition 5.9. Advice about an assumption:**  $(A, x, c, D)$  is an instance of advice-about-an-assumption if and only if (1)  $(A, x, c, D)$  is an instance of advice-about-a-cue and (2)  $c$  is an instance of assumption.

**Definition 5.10. Advice about a goal:**  $(A, x, c, D)$  is an instance of advice-about-a-goal if and only if (1)  $(A, x, c, D)$  is an instance of advice-about-a-cue and (2)  $c$  is an instance of goal.

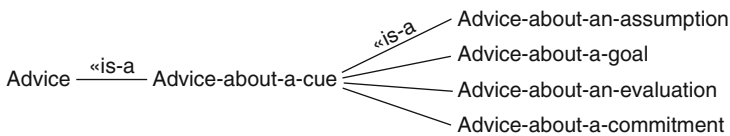
**Definition 5.11. Advice about an evaluation:**  $(A, x, c, D)$  is an instance of advice-about-an-assumption if and only if (1)  $(A, x, c, D)$  is an instance of advice-about-a-cue and (2)  $c$  is an instance of evaluation.

**Definition 5.12. Advice about a commitment:**  $(A, x, c, D)$  is an instance of advice-about-a-commitment if and only if (1)  $(A, x, c, D)$  is an instance of advice-about-a-cue and (2)  $c$  is an instance of commitment.

We have six concepts in the ontology of advice. Advice is the root concept, specialized into advice-about-a-cue, which is in turn specialized into four classes, depending on its target. We sought plausible targets in the decision information ontology, which gave us four concepts cited above. The result is a taxonomy of advice, illustrated in Fig. 5.3.

Figure 5.3 begs the question of whether all advice is advice about cues? Is every instance of advice also an instance of advice-about-a-cue?

One point of contention comes from discussions of intentions and commitments, which ask if the side-effects of intended actions are also intended. If a businessman intends to reduce costs by deciding to dump waste into a river instead of paying to dispense with it via recycling, then is the ensuing pollution of the river intended by



**Fig. 5.3** A taxonomy of advice. Any advice about a cue is either an about an assumption, a goal, an evaluation, or a commitment

that businessman? Arguments have been offered in favor of the view that it is never correct to say that a side-effect was the result of intention (Mele 2001). Bratman is not alone to argue for another view, that in some specific circumstances, side-effects can be intentionally brought about (Bratman 1987). The circumstances in question are that the individual, say **D**, “intentionally does *z* without intending to do *z*, as long as doing *z* is foreseen and **D** is willing to accept *z* as a consequence of **D**’s action” (Adams and Steadman 2004, p. 173). In other words, **D** intentionally produced the side-effect *z* if *z* is foreseeable and **D** accepts that *z* resulted from his actions. Whatever the position taken regarding this matter, it leads us here to the evident observation that commitment to bring about some state of affairs that is deemed desirable does not necessarily bring about specifically that state of affairs. There is obviously the inability to perfectly control the context in which actions are taken, and to fully determine the effects of these actions. Among the various effects of performed actions, some will be side-effects. It follows that an advisor of that businessman may have anticipated the side-effect of dumping waste, and may advise the businessman by targeting the beliefs of the latter. The ensuing question is where does the advice on side-effects fit in the taxonomy of advice – what concept in that taxonomy does the advice about side-effects instantiate? The answer is that advice on side-effects is an instance of a concept that specializes the advice concept in the taxonomy in Fig. 5.3 and not an instance of some other concept. Once we ask why the advisor **A** would advise the businessman **D** regarding waste disposal, we can see that any answer depends on what intentional states the advisor ascribes to the businessman. In one scenario, **A** believes that **D** does not know the side-effects of dumping waste; **A**’s advice will thus aim to bring to light to **D** the consequences that the latter may not have anticipated – **A**’s advice in this case targets **D**’s beliefs. In a second scenario, **A** will advise **D** because **A** believes that **D** desires to harm the environment by dumping waste – **A**’s advice here targets **D**’s desires, and **A** may advise **D** that such hopes counter ethical principles. In a third scenario, **A** believes that **D** values profit more than helping keep the environment clean – here, **A**’s advice targets **D**’s evaluations: **A** may try to convince **D** that this preference for profit over keeping the environment clean is not appropriate, because profit will suffer in the near future precisely because the environment was not kept clean (e.g., because the government may step in and hold the businessman and the company liable for damages caused by dumping waste). In the final, fourth scenario, **A** advises **D** because **A** believes that **D** intends to dump waste – **A**’s advice will target **D**’s commitments: **A** may tell **D** that he should not dump waste and offer whatever reason **A** thinks may persuade **D** to drop his intention to dump waste. As a summary of the four cases, remark that **A**’s knowledge of the side-effects of dumping waste may lead **A** to ascribe various intentional states to **D**, and thus advise **D** to influence what **D** may believe, desire, prefer, or intend with regards to waste disposal. Insofar as **A** gives advice because of the intentional states that **A** ascribes to **D**, and which **A** wishes to influence, the advice about side-effects is not a kind of advice that requires a new concept in the taxonomy of advice.



In general, to accept that an advisor **A** of a decision-maker **D** gives advice because of the intentional states that **A** ascribes to **D**, and which **A** wishes to influence, is to accept that any advice falls into one of the categories in the taxonomy of advice in Fig. 5.3.

## 5.6 Reinterpreting Advice

To the specialist-turned-generalist, i.e., the individual who faces a choice beyond his own specialty, the question of whether the advice he is given is relevant, clear, and so on, and ultimately, what consequences its acceptance or rejection has on his decision, is a question that bundles together two separate others: (1) *what is the advice referring to within the context of the specific decision situation?*; and (2) *what is the purpose of that advice in that context?*

Finding out what some advice refers to in a context of a given decision situation was the topic of Chap. 3. Advice there were signs, the purpose of which is to refer to concepts and objects. The reference relation is thus key, its formation being determined in the context in which the signs were used. How the signs are used, and how they fit the gaze of the individual upon himself and others will, via some imprecise mechanism of reference lead the individual to understand advice that he is given. To understand was thus to find reference between signs and available concepts and objects, themselves relating what the receiver of advice obtains via communication, along with his own past and present experiences and rationalizations of these experiences. The conceptual analysis of advice – in its role of one way in which to answer what advice refers to – asks precisely the questions of what is referring to what else. It asks what definitions stand behind the variously complex signs, and within what webs of assumptions, of ontological commitments, these definitions realize their purpose for the individual who advises, the one who communicates the signs. Conceptual analysis of advice seeks, in the simplest of terms, the background against which advice is engineered. The reason for doing so is terribly simple: whether to accept or reject advice is either approached as a coin-tossing exercise in which there really is no need to position advice against its background, or a matter of some thinking which is informed by a conception of the background of the advice of interest.

Establishing the background of advice – the context of reference and the ontological commitments through which we can make out the purpose of advice – would not be so difficult if it was merely a matter of connecting signs to some definite and universal concepts and objects, and establishing some definite ontological commitments. What makes conceptual analysis of advice tricky is the openness of reference and that the ontological commitments are identified mostly by speculation. Definite *and* universal referents are elusive, especially when a sign targets an abstraction, a concept. Referents established in some context of communication, and over the course of that communication can very well be deemed definite for all practical purposes of that specific communication. If one wants another to open the door, and

both of them see the door that is to be opened, then it would be far-fetched in normal conditions to debate whether the door they both see is a definite referent of the sign *that door* they use if English speakers, *cette porte* is speakers of French, and so on. They seem to know what they are talking about, and we should not deprive them of that ability by mere twist of argument.

Even when the referents are definite, they are not universal: they may be identifiable without doubt in a given context, yet in another context, the same sign can be used to refer to another referent, or the same referent can be referred to by a different sign. The sign *that door* does not *always* refer to that specific door the two are both looking at. This recalls the problem of indexicals, the reference of which is, so to speak, notoriously dependent on the context in which the indexical is used. Proper names and numbering standards are engineered solutions that civilization offers, naming people in the hope that even if identity is fluid, the physical person will still be identifiable for all practical administrative purposes by a name and a number. Even with names and numbers in the national registrars, the reference relations thus established are not universal for they last for a limited time (i.e., the referent is lost at death) and not all names are equally loved (hence the omnipresence of some). However, *that door* need of course not refer to any physical particular which separates two spaces, can be opened, and so on. Perhaps the speaker is using the term *that door* as a figure of speech, a metaphor. There are two compatible ways to pin down what the sign is in such a case referring to. One is to continue with the communication, so as to acquire more signs, some of them perhaps referring to definite properties of the referent. Another way, when communication cannot be pursued is to speculate from what is available in the context alone, and from what was previously communicated. In both of these methods the aim is to narrow down the number of potential plausible reference relations, in which the sign of interest participates. The difference between the two lies in the amount of information from which alternative reference relations are identified, and from which a choice is made to commit to some of the considered reference relations. Both methods cannot, however, go so far as to claim a definite and universal reference relation between a sign and some referent thereof, precisely because signs are dissociated from the referents for which they stand; the sign is itself a particular that is distinct from the particular or universal it is referring to. If not, reference could not be context-dependent, and we would conveniently enough have a language that knows of no open reference, or antireferents.

If the aim of conceptual analysis of advice is to determine the definite and universal concepts and objects to which the advice refers, any attempt to such an analysis is doomed from its very start. Its failure is, however, not complete if definite referents of advice are sought within a context, and no universal referents are chased. What is then the answer to the first question above, namely, what is the advice referring to within the context of the specific decision situation? The answer will, of course depend on the specifics of the decision situation in which the question is asked. Beyond the obvious, the answer that conceptual analysis offers to this question are reference relations, but not of the universal kind: they will instead point at best to definite referents in the context of the decision situation, at worst to antireferents.

The second question – *what is the purpose of some specific advice in a given context?* – becomes of interest especially after it is recognized that reference can be open even within and in spite of a given context, and this up to the point of targeting an antireferent. A sign, itself a particular is distinct from its referent, the latter a particular or a universal. Given the dissociation of the sign from its referent and the openness of reference, the problem the recipient is facing is not only if some advice is true or false. It is also what purpose that advice serves, what behavior it attempts to induce. When signs point to antireferents, evaluating the truth of signs is not a matter of determining whether the relations that signs state do correspond to relations between the objects and concepts the signs refer to. It is not only to determine whether the sign *that door is open* indeed reflects the actual state of affairs, in which that door over there is open. The question of purpose asks why it was said that *that door is open* in the first place.

Finding the purpose of some specific advice in a given context is a play in two acts. To start with, it is necessary to abandon hope in universal reference relations. Instead of searching to establish references once and for all, then redeploy these universal references to interpret advice in different decision situations, it is necessary to settle only with the possibility to find definite referents within some given context and even admit antireferents when no definite referent arises in the context. This is, in other words, to accept that the result of conceptual analysis is at best only good enough, appropriate for all practical purposes of the decision situation, in which the conceptual analysis of advice is applied. This chapter started off with the discussion of open reference, metaphor, and antireferents precisely to accomplish this first act. Openness of reference, the use of metaphor, and the possibility of antireferents all indicate that advice can be engineered with a purpose, and thus is given to advance that purpose. It was the aim in the second act in this chapter to look through the eyes of the advisor to see what advice may be used for. To the advisor of a decision-maker, advice is a tool through which the decision-maker can be influenced toward coordination. As the mind of the decision-maker is not immediately accessible as an open book is to the advisor, the best the advisor can do is construct explanations about the other's past, present, and future behavior. Such explanations, as the common use of language illustrates, consists of the advisor ascribing intentional states – beliefs, desires, intentions, and so on – to the decision-maker. To say that “the other does so and so because he believes this and desires that” is to use language to make explicit the intentional states ascribed to the other. It is from these explanations that the advisor decides whether and what advice to give to the decision-maker. It is reasonable to argue that the advisor will give advice because these explanations led him to anticipate some, in his view, undesirable future behavior of the decision-maker. It is, roughly speaking, because the advisor disagrees with what he expects the decision-maker will choose, that he provides advice in the first place. The advisor's aim is to influence the intentional states that he ascribed to the individual, to influence some specific element of the explanation the advisor had constructed for some prediction he made about the decision-maker. This of course means not that the decision-maker truly has “in his mind” these intentional states that the advisor ascribed. Rather, the intentional states are merely a tool that the

observer-advisor uses to organize his own explanations and predictions of the past, present, and future choices of the decision-maker whom the advisor is observing and advising.

The general answer to what purpose advice has is that an advisor engineers and dispenses advice to influence the choice of decision-makers. The advisor ascribes intentional states, and on this basis predicts the choices of the decision-maker; from there on, the advisor determines whether he will accord with the anticipated choices, and if not, engineer and give advice in order to influence these choices. Advice is thus engineered on the basis of speculative explanations and predictions, the relevance of which will ultimately depend on the advisor's experience of the decision situation and of his past interactions with the decision-maker whom he is advising.



## Chapter 6

# Advisor's Problem and Its Solutions

An engineered ontology is an essential part of any set of assertions produced toward explanation and prediction. It is an integral part of *any* theory, regardless of it being scientific. To perform its explanatory and predictive roles, a theory must include a language through which it points to, refers to universals and particulars, concepts and phenomena. The language of classical mechanics speaks of time, distance, mass, and force, while that of game theory talks of players, games, and strategies. The terms essential to the explanations and predictions of a theory form the engineered ontology that underlies the theory.

When an exercise in conceptual analysis ends in discontent, ontology engineering ensues. To be dissatisfied with one set of concepts comes with the question of what other available or potential set may better suit the primarily explanatory and perhaps some predictive purposes of a theory we are concerned with. It is a change of theory that is attempted, the aim being to replace one language with another with the hope that the latter is more appropriate to the task of interest. To ask what conditions any candidate language, any set of concepts should satisfy is essentially a question of how well the alternative languages help in explanation and prediction, and ultimately – and especially if the aim is to produce a scientific theory – how they fit experience and the outcomes of systematic and reproducible empirical observation.

If a conceptual schema organizes experience toward explanation, an engineered ontology is merely an attempt at giving an explicit account, at laying out the signs referring to the elements of the conceptual schema. An engineered ontology is thus but a collection of signs having the distinctive trait that if engineered carefully enough, the referents of the signs will be concepts and objects that have a role in the construction of explanations and predictions which will ideally be agreeable to more than a single individual. When the endeavor is successful, when the ontology underlies a theory that produces explanations and predictions acceptable to many and/or when alternatives seem less capable, the theory is perpetuated. The theory of classical mechanics, for example, is taught already for quite some time for it plays exceptionally well its roles within the limits of speed and size which separate it from quantum and relativistic mechanics.

Theories such as that of classical mechanics are engineered from the conclusions of rigorous and longtime enquiry in which the link between the concepts and phenomena is accessible to anyone interested enough to learn that language and use

it for explanation and prediction that she can experience herself. It is obvious, but still appropriate, to say for such theories that they do reflect knowledge accumulated over time and continually subjected to criticism.

Not all theories are or can be engineered as rigorously and over as much time as stable scientific theories. Facing a decision-maker, the advisor forms explanations and predictions of the other's behavior, for it is from such grounds that the advisor will choose the form and content of advice. Explanations and predictions of other's behavior proceed – following our earlier argument – through the ascription of intentional states: the advisor forms a theory about the decision-maker's behavior by ascribing intentional states to her. In doing so, the advisor engineers a theory which will then serve to explain and predict the decision-maker's behavior.

The theory of, again, classical mechanics and a theory that an advisor makes about the decision-maker before giving advice are only really similar in two respects. They both (1) aim to explain, and perhaps predict, and (2) should be subject to the same basic rule, which is if they fail in either explaining and predicting, they should be revised and/or their scope of applicability restricted. The two otherwise stand in stark contrast. The former is the result of longtime systematic enquiry, while the latter comes out of reflection in a given context and under the very severe constraints of time and attention, among many other rare resources. The former theory is very general, in the sense that it can be applied in a variety of settings; e.g., laws of classical mechanics apply as long as the dimensions are far above  $10^{-9}$  m and speeds remain far below the speed of light. In other words, laws of classical mechanics are good enough for explanation and prediction of the motion of bodies in most situations in which humans tend to find themselves in, given the current technology and its foreseeable future. If the axioms of perfect rationality were laws on par to those of classical mechanics, then an advisor could assume that the decision-maker is perfectly rational, and as we argued earlier, should not advise her at all – there would really be no need, for the decision-maker already knows all that she needs to know. Axioms of perfect rationality and of any variant of bounded rationality are far from being anywhere near the laws of classical mechanics. It is unclear under what practical conditions we can assume that the regularities formalized via these axioms do in fact apply, and thus relevantly describe the behavior to expect of the decision-maker.

Every advisor has the misfortune to lack a general and empirically valid theory of choice, which she would apply to explain and predict the behavior of those to advise. The absence of a general theory certainly does not mean that the advisor is somehow paralyzed when she should explain and predict the behavior of the recipients of advice, and subsequently choose what advice to dispense. Instead of relying on a general theory, the advisor's explanations and predictions arise out of much less general theories, but theories nevertheless, thus amenable to description and testing. Such a theory does get implicitly described anyway, for the decision-maker must somehow organize and draw conclusions from the cues that she has about the decision-maker. The advisor also tests her theory, for the explications and predictions it produces can be contrasted to the decision-maker's past and present behaviors, those which the advisor can observe.

To the extent that an advisor's theory about a decision-maker is not a general theory, but quite literally one formed on the basis of little, in rare cases systematically collected evidence and argument, and which has a very limited scope of application, we will call such a theory a *picotheory*. Engineered on the basis of limited information – certainly limited in contrast to a scientific theory – a picotheory articulates the various cues that an advisor has about the behavior of the individual whom she is to advise. A picotheory is both specific to the context in which the advisor should dispense advice, and thus obviously specific to the person or people who will receive advice. Instead of thinking that an advisor will base her explanations and predictions of the decision-maker's behavior on some universal theory that is in effect still elusive, a picotheory reflects that the advisor will instead put together some highly context- and person-specific account of why the other behaves in some way and how that other will end up choosing. The resulting picotheory serves to the advisor as a model of the decision-maker. A picotheory will thus certainly be made of that which the advisor ascribes to the decision-maker, i.e., the cues.

What we call the *advisor's problem* is essentially this: given an advisor's explanations and predictions of a decision-maker, what is the advice that this advisor should give to that decision-maker to influence the predicted choice of the decision-maker? An answer is sought in three steps. First, the advisor forms a picotheory to predict what the decision-maker will choose. Given the explanation and prediction of the picotheory, the advisor should see what changes to make in the picotheory – i.e., revise it – for it to predict the choice more desirable to the advisor. Finally, based on the adjustments made in the second step, the advisor should determine the content and form of advice that he should give to the decision-maker so that the latter chooses what the revised picotheory predicts. Basically, one starts by predicting to form a picotheory, then revises this initial picotheory so that it predicts that the decision-maker will choose what the advisor prefers, and finally defines advice that would reorient the choice toward the one predicted in the revised picotheory.

Picotheories are obviously unstable constructions, to be revised whenever they fail to explain or anticipate properly the behavior of the decision-maker. Being specific to particular decision-makers and contexts of choice, picotheories certainly are to a considerable extent speculations that can never be as rigorous and as stable as laws established over long periods through structured and rigorous theorizing and verified via controlled experiment, and this will remain so as long as empirically valid laws that govern people's choices remain elusive.

To the extent that universal laws governing human behavior are at present unknown as much as is necessary to produce an empirically valid model of choice, any attempt at the engineering of advice involves the elaboration and revision of picotheories. The aim of this chapter is to discuss what may go into an advisor's picotheory and how a picotheory may be revised in the presence of new information, such as that obtained via communication, or the advisor's observation of the decision-maker's behavior. Once we have said below what picotheories may look like, we will be able to offer a finer definition of the advisor's problem toward the end of this chapter.



## 6.1 Advice from Simple Explanations and Predictions

A picotheory is an advisor's context- and decision-maker-specific theory about why and how a decision-maker behaves in the given context. It is thus a model of choice that an advisor defines before she considers what advice she should give to the decision-maker. As a model of choice, a picotheory will be a poor one, but good enough precisely because not much better can be done by the advisor who is, just as Herbert Simon's administrative man, an individual both of limited memory and reasoning abilities. Worse even, the advisor must, within such limits, explain and predict the decision-maker's behavior. Otherwise, the advice that is offered must be irrelevant, except when sheer luck interferes. A picotheory has four parts:

1. A decision information ontology to classify information that forms explanations and predictions about the decision-maker
2. Instances of concepts from the decision information ontology
3. A decision problem to explain how the content gives a picture of the decision situation
4. Decision rules to state the properties of candidate solutions.

### Decision Information Ontology: Ontology of Cues

The advisor's theory will need a set of concepts, an ontology that organizes the information is relevant to the decision-maker. It is by articulating that information, the advisor believes, that the decision-maker decides to act. The advisor here faces the question of what information in a given decision situation is relevant not from her own perspective, but from that of the decision-maker whom she is advising. A decision information ontology gives an answer. Insofar as the advisor ascribes intentional states to the decision-maker, the decision information ontology of cues says that, in the eyes of the advisor, the decision-maker reasons on the basis of some beliefs, desires, intentions, evaluations, and forms commitments to act in some way. We argued in Chap. 4 that there are good reasons why the ontology of cues identifies the principal kinds of information used in decision making, and it will be assumed in this chapter that the ontology of cues is the decision information ontology used in picotheories.

### Instances of Cues

To have the ontology of cues as the decision information ontology in picotheories is to say that the advisor explains and predicts a decision-maker's choices from the intentional states the former ascribes to the latter. Now, it has been argued – as we discussed in relation to speech acts – that there are two intertwined parts to any intentional state: its content and its psychological mode. A decision information ontology, the first part of the advisor's theory, classifies decision information on

the basis of psychological mode, and hence, there are the concepts of assumption, goal, evaluation, and commitment. But that ontology says nothing about the content of intentional states. Yet, when an advisor ascribes intentional states, she must ascribe both the psychological mode and the content. The point is that by ascribing intentional states, an advisor makes two ascriptions: (1) she ascribes a conceptual schema to the decision-maker, and (2) she ascribes psychological modes about the elements of that conceptual schema. Consider an example – suppose that the advisor **A** believes that the decision-maker **D** desires to live in a totalitarian society. There are two parts to what **A** ascribes here to **D** : one is the property “living in a totalitarian society” and the other is that **D** desires to exhibit that property. Two questions ensue: (a) is it appropriate to ascribe that property (i.e., content of intentional state) to **D** ? and (b) is it appropriate to ascribe that psychological mode to **D** in relation to that property? Conceptual analysis answers the first question: the purpose of the analysis is to see what **D** understands by “totalitarian society,” i.e., what this phrase refers to according to **D**, and how close or distant this understanding is from the one that **A** has of the same phrase. Conceptual analysis answers thus the first question by looking into the differences between what “totalitarian society” refers to according to **A** and **D** . If conceptual analysis leads **A** to believe that there are few similarities between the references she and **D** have for the phrase “totalitarian society,” then it is relevant for **A** to revise the intentional state that she ascribed initially to **D** . The second question – whether the ascribed psychological mode is appropriate – will be informed by conceptual analysis as well: e.g., if conceptual analysis indicates that the phrase “totalitarian society” has for **D** a reference which **D** evaluates negatively, then it is perhaps questionable for **A** to believe that **D** desires to live in a totalitarian society if **D** seems to despise it. To answer both questions, conceptual analysis needs to rely on the advisor’s conclusions from her prior and present communication and interaction with the decision-maker, along with her grasp of the context in which the decision is being made.

## Decision Problem

Neither explanation nor prediction can be done only by ascribing intentional states. Doing so provides the input to any attempt to explanation and prediction. What is necessary then is a process that transforms these inputs into some outputs of interest. Recall the division in a model of choice: any model of choice has three parts, the decision information ontology, the decision problem, and the decision rules. The decision information ontology tells the decision-maker how to classify the information available to her. The decision problem says how the decision-maker should relate the different kinds of decision information (i.e., the instances of the concepts in the decision information ontology) to gain a picture, a conceptualization of her decision situation. Finally, decision rules define properties that a solution to the decision problem should have.

By ascribing intentional states, the advisor introduces in the picotheory the inputs, but still lacks the decision problem which would organize these inputs. To

organize the inputs is basically to establish some relations between them. The decision problem in a picotheory plays the same role as the decision problem in a model of choice. Suppose that an advisor **A** has ascribed to **D** some intentional states, that **A** thus has some cues about **D**. According to the ontology of cues, the advisor will have four sets of cues: a set of assumptions, another of goals, a third set of evaluations, and a fourth that carries commitments. These four sets are the inputs that a decision problem in a picotheory will need to organize. Suppose that **A** believes that **D** will try to satisfy the goals that **D** evaluates as most desirable (i.e., the goals she prefers most), all the while not violating the assumptions **D** holds. In other words, **D** wishes to satisfy desires while not violating beliefs. This hypothesis that **A** makes immediately gives a decision problem formulation for **A**'s picotheory about **D**. The decision problem is then this: Given a set of assumptions, goals, and evaluations, the decision-maker needs to identify the commitments that lead her to satisfy the goals she prefers most and that will not violate the assumptions. According to this decision problem, the picotheory says that the decision-maker will seek commitments.

### Decision Rules

Decision rules are properties that a choice must satisfy. Decision rules are closely related to the decision problem formulation in a picotheory. Suppose, as in the above paragraph, that the problem of the decision-maker is to identify the commitments that lead her to satisfy the goals she prefers most and that will not violate the assumptions. In a picotheory that incorporates this as the decision problem, the following two are the obvious decision rules: (1) the chosen commitments must lead the decision-maker to satisfy the goals she prefers most, and (2) the chosen commitments should not lead the decision-maker to violate assumptions.

### The Four Together

How do these four elements come together into a picotheory that can explain and predict? It is clear that subtracting any of the four leaves us something of little use. If we eliminate for example the instances, we have an empty picotheory, in the sense that it is context-independent, and that it is unclear what the decision situation is about anyway. If we eliminate the decision rules, we cannot predict anything: any potential choice of the decision-maker will be equally good, since there are no criteria that a choice should satisfy in the absence of decision rules. If the advisor makes no assumptions about what decision rules the decision-maker aims to satisfy, the advisor's prediction is no better than flipping a coin.

A picotheory has four parts. Two of these concern the inputs, the information from which speculative explanation and prediction arise. The two are the decision information ontology, which classifies the psychological modes of the intentional states that the advisor ascribes to the decision-maker. The other of these two are the instances, which reflect the conceptual schema behind the content of the intentional

states that the advisor ascribed to the decision-maker. It is, to recall the argument from Chap. 4, the observer who ascribes intentional states to the observed individual; in a situation in which advice is given, the former is the advisor, the latter the decision-maker being advised. By performing the ascription, the advisor forms the first part of her picotheory, the purpose of which is to explain and predict the behavior of the decision-maker. Both explaining and predicting are necessary to the advisor before she chooses the advice to dispense, for the very purpose of advice is to influence the initially predicted behavior, to orient it toward behavior which may be more desirable in the eyes of the advisor. To ascribe thus intentional states is to hypothesize about the psychological moods of the decision-maker and about the conceptual schema through which she experiences the decision situation. The third part of the picotheory takes the elements of the ascribed intentional states, that is, takes these cues and relates them into a decision problem. The decision problem is a description of what challenge the decision-maker faces, from the perspective of the advisor. The three still cannot do much without the final, fourth part of a picotheory. The decision rules will say which responses to the challenge are desired by the decision-maker, and this again, from the eyes of the advisor. It is the decision problem and rules that say how the decision-maker will behave, given some intentional states that the advisor ascribed to her.

## 6.2 Models of Advice: An Overview

Conceptual analysis discussed in the preceding chapters led us to define and specialize cue and advice concepts. Moving toward these definitions illustrated how conceptual analysis can be performed and what its result can be. Any picotheory, a structure made of a decision ontology, its instances, of a decision problem and rules can readily be subjected to conceptual analysis. Questions can certainly be asked of what reference relations arise from the elements of the picotheory.

Insight obtained by conceptual analysis is certainly not all that we can and should explore when analyzing or engineering advice. Further conclusions about what to advise can be drawn by looking into the logical structure of picotheories. A study of the logical structure does not, roughly speaking, focus on reference relations and hence interpretation of signs or their combinations, but on relations between combinations of signs and the properties of these relations.

### 6.2.1 Introductory Example

*Example 6.1.* To grasp intuitively what is meant by the “structure” of a picotheory, consider a trivial example. An advisor **A** is looking at a decision-maker **D**. What **A** sees is that **D** is standing in front of a door, and the door is closed. **A** believes that **D** wants to open the door and **A** knows that the door is locked. **A** has the key that

can unlock the door. Finally, suppose that **A** desires to help **D** to open the door. If **A** reflects on this situation, he might conclude that he should give that key to **D**. We can rewrite this case as follows:

- $x_1$ : it looks to **A** that **D** wants to open the door
- $x_2$ : it looks to **A** that **D** does not have the key
- $x_3$ : it looks to **A** that **D** knows that the door is locked
- $x_4$ : **A** knows that the door is locked
- $x_5$ : **A** has the key that can unlock the door
- $x_6$ : **A** desires that **D** opens the door
- $x_7$ : if **A** knows that the door is locked and **A** has the key that can unlock the door and **A** desires that **D** opens the door, then **A** gives the key to **D**
- $x_8$ : if **A** gives the key to **D**, then **D** has the key that opens the door
- $x_9$ : it looks to **A** that if **D** has the key that opens the door and assumes that the key can be used to open the door, then **D** commits to open the door
- $x_{10}$ : it looks to **A** that if **D** commits to open the door, then the door will be opened

The ten sentences above give us a description of the situation. Each sentence is preceded by the indexed letter  $x$ . The choice of the letter  $x$  is arbitrary, and indexes are given in increasing order as a matter of notational convention. Just as **D** is an abbreviation of the decision-maker, so does an indexed  $x$ , e.g.,  $x_1$  abbreviate a natural language sentence written to the right of  $x_1$ .

We can obviously take every sentence above, or any set of these individual sentences and be concerned with their interpretation by applying conceptual analysis. This might lead us to ask if there is more than one door, which of them does the key unlock, do **A** and **D** focus on the same door, if the door is a metaphor for something, and so on.

To be interested in the logical structure of the information that describes the situation above is to ask what conclusions can be drawn from the given sentences and study the properties of those conclusions. **A** may conclude from the above that he should give the key to **D**, so that **D** will have the key, and will open the door because it looks to **A** that **D** wants to open the door. Although this sounds fine, it would be considerably more interesting to be able to write down the precise conditions which this conclusion satisfies, so as to decide whether it is a valid one. To do so, we first need to rewrite complex sentences according to some simple rules. The purpose of these rules is to break down complex sentences and surface bits and pieces that satisfy some specific conditions.

The first rule is to distinguish, as in Chap. 3 the psychological mode from the content. We do this by taking out verbs for psychological modes from the sentences, while rewriting the content as simple declarative sentences of English. Following Definitions 5.1–5.12, and given the psychological mode and proposition in a sentence, we can say which concept is instantiated by the sentence. In  $x_1$ , **A** is ascribing the desire “the door is open” to **D**. We can consequently rewrite  $x_1$  as the cue  $\mathbf{g}p_1$ , where  $\mathbf{g}$  says us that  $p_1$  is an instance of the goal concept, while  $p_1$  abbreviates the proposition “the door is open.” We can thus first extract the propositions in  $x_1, \dots, x_6$ :

- $p_1$ : the door is open
- $p_2$ : the key is missing
- $p_3$ : the door is locked
- $p_4$ : a key is available
- $p_5$ : **D** opens the door

The first rule leads to the instantiation of the CUE concept, according to the Definitions 5.2–5.5.

To understand the second rule, remark that the only difference between  $x_3$  and  $x_4$  is in which of the individuals holds or ascribes the proposition  $p_3$ . It is thus useful not only to label a proposition by a symbol for the concept it instantiates, but to indicate also the role of the individual whose intentional states are about the proposition. We have only two roles, the advisor and the decision-maker being advised. This leads to the following rules that serve to label propositions:

1. If a proposition  $p$  is held by the advisor and
  - (a)  $p$  is an instance of assumption, then write it **a<sub>A</sub>** $p$
  - (b)  $p$  is an instance of goal, then write it **g<sub>A</sub>** $p$
  - (c)  $p$  is an instance of evaluation, then write it **e<sub>A</sub>** $p$
  - (d)  $p$  is an instance of commitment, then write it **c<sub>A</sub>** $p$
2. If the advisor ascribes a proposition  $p$  to the decision-maker and
  - (a)  $p$  is an instance of assumption, then write it **a<sub>D</sub>** $p$
  - (b)  $p$  is an instance of goal, then write it **g<sub>D</sub>** $p$
  - (c)  $p$  is an instance of evaluation, then write it **e<sub>D</sub>** $p$
  - (d)  $p$  is an instance of commitment, then write it **c<sub>D</sub>** $p$

Using the rules above, we can rewrite  $x_1, \dots, x_{10}$  as shown below:

$$\begin{array}{ll}
 x_1 \equiv \mathbf{g}_D p_1 & x_6 \equiv \mathbf{g}_A p_5 \\
 x_2 \equiv \mathbf{a}_D p_2 & x_7 \equiv \mathbf{a}_A p_3 \wedge \mathbf{a}_A p_4 \wedge \mathbf{g}_A p_5 \rightarrow \mathbf{c}_A p_6 \\
 x_3 \equiv \mathbf{a}_D p_3 & x_8 \equiv \mathbf{c}_A p_6 \rightarrow \mathbf{a}_D p_4 \\
 x_4 \equiv \mathbf{a}_A p_3 & x_9 \equiv \mathbf{a}_D p_4 \wedge \mathbf{a}_D p_7 \rightarrow \mathbf{c}_D p_8 \\
 x_5 \equiv \mathbf{a}_A p_4 & x_{10} \equiv \mathbf{c}_D p_8 \rightarrow \mathbf{g}_D p_1
 \end{array}$$

The purpose of the third rule is to allow us to look into the internal structure of sentences that contain potentially complicated combinations of psychological modes and simple propositions. This was not necessary for  $x_1, \dots, x_6$ , but certainly is for  $x_7, x_8, x_9$ , and  $x_{10}$ . We break up these complex sentences in two ways.<sup>1</sup> One

<sup>1</sup> There are different ways to extract relations from sentences, and how it happens depends essentially on what relations one wishes to highlight. The more the relations, the more likely is the formalism that manipulates these relations to be complicated. The point here is to remain in a very simple case, hence the few relations that we take into consideration.

way is to have the conjunction relation, i.e., *and*, outside of a sentence which states a conjunction of simpler propositions. The other is to have the *if-then* relation outside of a sentence in which this conditional relation appears. This is essentially a rewriting of complex sentences so that these *and* and *if-then* relations do not remain hidden in the natural language presentation of the complex propositions. We will use the symbol  $\wedge$  to refer to the *and* relation, while  $\rightarrow$  refers to the *if-then* relation.

Using the above three rules, we can rewrite  $x_7$ ,  $x_8$ ,  $x_9$ , and  $x_{10}$  as follows:

- $x_7$  is rewritten as  $\mathbf{a_A}p_3 \wedge \mathbf{a_A}p_4 \wedge \mathbf{g_A}p_5 \rightarrow \mathbf{c_A}p_6$ , where  $p_6$  refers to the proposition “give the key that opens the door to **D**”
- $x_8$  as  $\mathbf{c_A}p_6 \rightarrow \mathbf{a_D}p_4$
- $x_9$  as  $\mathbf{a_D}p_4 \wedge \mathbf{a_D}p_7 \rightarrow \mathbf{c_D}p_8$ , where:
  - $p_7$  refers to “the key can be used to open the door”
  - $p_8$  refers to “open the door”
- $x_{10}$  as  $\mathbf{c_D}p_8 \rightarrow \mathbf{g_D}p_1$

There are two consequences to having relations “outside of” sentences. First, we can distinguish simple from complex sentences, the former being those in which there are neither *and* nor *if-then* relations between simple sentences. A proposition is a sentence without either *and* or *if-then*. Second, we can now introduce rules of deduction, or equivalently, rules of inference. Such rules let us draw conclusions from some given set of simple and complex propositions. For example, if **A** has  $\mathbf{a_A}p_3$ ,  $\mathbf{a_A}p_4$ ,  $\mathbf{g_A}p_5$ , along with  $\mathbf{a_A}p_3 \wedge \mathbf{a_A}p_4 \wedge \mathbf{g_A}p_5 \rightarrow \mathbf{c_A}p_6$ , then **A** can conclude  $\mathbf{c_A}p_6$ . Reaching that conclusion can be represented as an application of an inference rule, i.e., by considering that **A** uses the inference rule called *modus ponens* which has the following general form, for any (natural number)  $n \geq 1$  and any set of simple or complex sentences  $\{y_1, \dots, y_n, z\}$ :

$$\frac{y_1 \quad y_2 \quad \dots \quad y_n \quad y_1 \wedge y_2 \wedge \dots \wedge y_n \rightarrow z}{z}.$$

Above the line are the *premises* of the rule application, below it is its *conclusion*. The application of the rule refers to a step of inference, of reasoning that was performed. To conclude that **D** opens the door in this example, we consider that **A** applies modus ponens three times, first to conclude **A**'s commitment to give the key that opens the door to **D**, i.e., to conclude  $\mathbf{c_A}p_6$ :

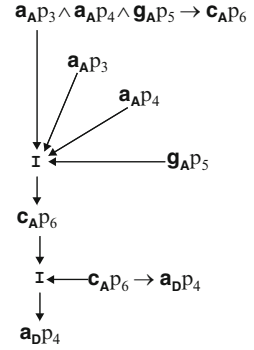
$$\frac{\mathbf{a_A}p_3 \quad \mathbf{a_A}p_4 \quad \mathbf{g_A}p_5 \quad \mathbf{a_A}p_3 \wedge \mathbf{a_A}p_4 \wedge \mathbf{g_A}p_5 \rightarrow \mathbf{c_A}p_6}{\mathbf{c_A}p_6}.$$

Once we have  $\mathbf{c_A}p_6$  and  $\mathbf{c_A}p_6 \rightarrow \mathbf{a_D}p_4$ , we can conclude  $\mathbf{a_D}p_4$ :

$$\frac{\mathbf{c_A}p_6 \quad \mathbf{c_A}p_6 \rightarrow \mathbf{a_D}p_4}{\mathbf{a_D}p_4}.$$

The advisor at this point concludes that if he gives the key to the decision-maker, the latter will have a key. We can represent the advisor's reasoning up to

**Fig. 6.1** A representation of the advisor's reasoning up to his concluding  $\mathbf{a_D}p_4$ , according to Example 6.1. Each node is either a labeled proposition, or is “ $\mathbf{I}$ ” which refers to the application of modus ponens



him concluding  $\mathbf{a_D}p_4$  as a graph in Fig. 6.1. Every node in Fig. 6.1 is either a cue or is “ $\mathbf{I}$ ” which refers to an application of modus ponens. The graph was drawn simply by connecting the premises of a modus ponens application to an  $\mathbf{I}$  node, then connecting that node to the conclusion of that same modus ponens application.

According to  $x_9$ , which was reformulated as  $\mathbf{a_D}p_4 \wedge \mathbf{a_D}p_7 \rightarrow \mathbf{c_D}p_8$ , the advisor assumes that if  $\mathbf{D}$  has the key that opens the door and assumes that the key can be used to open the door, then  $\mathbf{D}$  commits to open the door. Given  $\mathbf{a_D}p_4$  and  $\mathbf{a_D}p_4 \wedge \mathbf{a_D}p_7 \rightarrow \mathbf{c_D}p_8$ , the advisor cannot conclude that the decision-maker would open the door. Here, the advisor can either assume that  $\mathbf{a_D}p_7$ , i.e., that  $\mathbf{D}$  believes that the key she has can open the door, or the advisor can assume that  $\mathbf{D}$  believes no such thing. In the former case,  $\mathbf{a_D}p_7$  is a cue, while in the latter case, the advisor should produce advice that supports  $\mathbf{a_D}p_7$ . If we take the latter instead of the former case, this leads to a puzzle for the advisor, illustrated in Fig. 6.2. What is shown are advisor's conclusions up to  $\mathbf{a_D}p_4$ , after which the advisor cannot conclude immediately that  $\mathbf{g_D}p_1$ : the reason why this is not a valid conclusion is that  $\mathbf{a_D}p_7$  is missing. The failed inferences are shown using dashed lines in Fig. 6.2.

In order to conclude  $\mathbf{g_D}p_1$ , the advisor ought also to give the advice to support the proposition that the decision-maker can use the key to open the door. Assuming then that  $\mathbf{A}$  advises that  $\mathbf{D}$  should assume that  $p_7$ , the advisor can conclude  $\mathbf{c_D}p_8$ :

$$\frac{\mathbf{a_D}p_4 \quad \mathbf{a_D}p_7 \quad \mathbf{a_D}p_4 \wedge \mathbf{a_D}p_7 \rightarrow \mathbf{c_D}p_8}{\mathbf{c_D}p_8}.$$

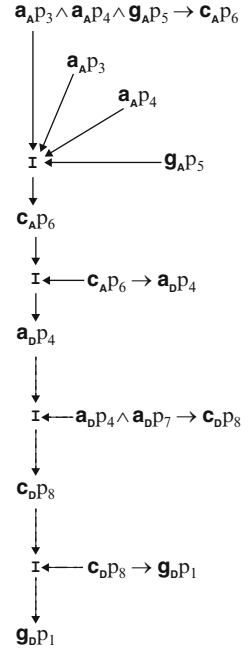
Finally, as  $\mathbf{D}$  commits to open the door and  $\mathbf{c_D}p_8 \rightarrow \mathbf{g_D}p_1$ , the advisor can conclude that the decision-maker's goal is satisfied:

$$\frac{\mathbf{c_D}p_8 \quad \mathbf{c_D}p_8 \rightarrow \mathbf{g_D}p_1}{\mathbf{g_D}p_1}.$$

We will not go in this example into the details of why  $\mathbf{A}$  should advise specifically  $\mathbf{a_D}p_7$  to  $\mathbf{D}$  and not suggest also or instead something else to  $\mathbf{D}$ . This is a question of what to advise so as to avoid advising too much or too little, hence being a question of efficiency which we will discuss in detail later in this chapter.



**Fig. 6.2** An illustration of how reasoning may move from  $\mathbf{a_D}p_4$  using the information available to the advisor. Lines are *dashed* to show that inference fails, that is, cannot be performed as a premise is missing. The figure thus shows a puzzle for the advisor, as he should provide advice that supports  $\mathbf{a_D}p_7$ . According to the cues that the advisor has about the decision-maker, only  $\mathbf{a_D}p_7$  is missing for the advisor to conclude that  $\mathbf{g_D}p_1$



To synthesize, we started from natural language sentences  $x_1, \dots, x_{10}$  describing a situation in which the decision-maker **D** has to choose how to act and an advisor **A** is present. From **A**'s perspective, **A** and **D** have different information at their disposal, modeled via the different assumptions, goals, and so on. **A** is thus interested in **D**'s goals, assumptions, commitments, and evaluations, and how they relate to **A**'s own. We rewrote  $x_1, \dots, x_{10}$  so as to surface the propositions therein, the roles these propositions are attached, as well as their classification among the extensions of the concepts that specialize cue. A picotheory in this example are all the labeled propositions, all the inference rule applications, advice given, and conclusions drawn – stated otherwise, the entire example.

The rest of this section will go over the rules and the mathematical devices introduced informally in Example 6.1. This will result in a framework for the representation and reasoning about picotheories, or equivalently, models of advice. The mathematics are elementary in the sense that no knowledge of advanced topics is needed, and all that is needed is presented in this chapter. Emphasis is on simplicity so as to surface core ideas. Beyond facilitating presentation, models of advice are interesting by being an important part of any rigorous approach to the engineering of advice. If the reader is persistent in this section, the reward is a more thorough and clearer understanding of the discussions of the advisor's problem, decision rules, of the resolution of the problem, and thereby of models of advice and their use.

### 6.2.2 Overview of the Framework for the Modeling of Advice

To work out the introductory example (Sect. 6.2.1), it was necessary to perform three tasks: (1) the *classification* of information, (2) its *representation* in models where labeled propositions (e.g.,  $\mathbf{C_A} p$ ) and formulas (e.g.,  $\mathbf{a_D} p_4 \wedge \mathbf{a_D} p_7 \rightarrow \mathbf{C_D} p_8$ ) are related via inference relations (among others, as we shall see below) and perhaps visualized as graphs, and (3) some form of *analysis*, such as determining what to advise the decision-maker. These tasks were performed in a particular way, in the sense that the information relevant to the example was categorized by following the ontologies of cues and advice. The resulting instances of cue and advice were represented in a particular way as well, for only one kind of inference rule was used (i.e., modus ponens) and formulas highlighted only the conjunction and if-then relations. Analysis was rather poor, as it only involved looking into what cue was missing, so as to point out what the decision-maker should be advised.

A framework here refers to a construction made of three parts. One is a scheme for the classification of information, that is, one ontology, or more of them used together. The second component are facilities for modeling – a mathematical formalism at least – to record, represent, and relate in models the instances of the concepts of the ontologies that classify. The first two components are designed to enable the modeling of information relevant for the analysis of advice, that is, to capture cues and advice. It is through the construction of a model of advice, through classification and modeling that the advisor's problem is defined in a particular advice(-giving or -taking) situation. The third and final part are methods, rigorous and organized guidelines for what questions to ask a model of advice and how to obtain answers thereto. A framework thus serves to write and study picotheories: classification amounts to the instantiation of one decision information ontology (or more of them used together), modeling involves the representation and relation of the instances of the decision information ontology, while analysis encompasses the definition of the decision problem from the constructed models, as well as the application of decision rules so as to identify candidate solutions to the decision problem.

The framework for the analysis of advice uses the ontologies of cue and advice to perform the classification of information. Instances thereby obtained are represented in *advice networks*, graphs which capture advisor's inferences over cues and advice, conflicts that he may identify among cues and advice, and his evaluations of cues and advice. An advice network – an *a-net* – defines the problem that the advisor is facing when determining what to advise the decision-maker. Candidate solutions are sought either by transforming an a-net or by looking inside the a-net itself and searching for parts thereof which satisfy some specific conditions. The aim below is to go over the intuitive ideas about the framework for the analysis of advice, before moving to its mathematical formalization.

The advisor relies on own beliefs, desires, intentions, and evaluations about the decision situation and the decision-maker to determine what to advise. The relevant information may arise out of the advisor's experience, or otherwise from the current observation and interpretation of the decision-maker's communication, actions, and



two parts of the sentence:  $\mathbf{a_A}p_1$  = “the advisor believes that the door is open,”  $\mathbf{a_A}p_2$  = “the advisor believes that he has the key.”

- The **implication** relation rewrites the if–then structure of a sentence in a natural language. A sentence “**if** the advisor has the key, **then** he can open the door” can thus be abbreviated by  $\mathbf{a_A}p_1 \rightarrow \mathbf{a_A}p_2$ , where  $\rightarrow$  refers to the implication relation and reads “implies” (or if preferred, “if  $\mathbf{a_A}p_1$  then  $\mathbf{a_A}p_2$ ”), while the labeled propositions read  $\mathbf{a_A}p_1$  = “the advisor has the key” and  $\mathbf{a_A}p_2$  = “the advisor can open the door.”
- The **inference** relation refers to a step of reasoning, that is, the move by which one reaches a conclusion from a given set of premises. The inference relation in this framework refers to an application of *modus ponens*, which has the following form:

$$\frac{y_1 \quad y_2 \quad \dots \quad y_n \quad y_1 \wedge y_2 \wedge \dots \wedge y_n \rightarrow z}{z}.$$

Modus ponens relates a set (above:  $\{y_1, y_2, \dots, y_n, y_1 \wedge y_2 \wedge \dots \wedge y_n \rightarrow z\}$ ) of cues and advice to a single cue or advice (above:  $z$ ).<sup>2</sup>

- The **conflict** relation stands between cues and/or advice which contradict one another. Conflict is modeled as a step of inference which concludes logical inconsistency. Let the sign  $\perp$  refer to logical inconsistency, and suppose that  $\mathbf{a_A}p_1$  = “the advisor believes that he can open the door” and  $\mathbf{a_A}p_2$  = “the advisor believes that he cannot open the door.” To state that  $\mathbf{a_A}p_1$  contradicts  $\mathbf{a_A}p_2$ , that is, to introduce a conflict relation between them, the assumption that they indeed are contradictory is added:  $\mathbf{a_A}p_1 \wedge \mathbf{a_A}p_2 \rightarrow \perp$ . Given thus  $\mathbf{a_A}p_1$ ,  $\mathbf{a_A}p_2$  and  $\mathbf{a_A}p_1 \wedge \mathbf{a_A}p_2 \rightarrow \perp$  as premises of modus ponens, the conclusion is logical inconsistency,  $\perp$ .
- As not all cues and advice are equally desirable, the **preference** relation compares pairs of cues and/or advice in terms of desirability. A preference thus refers to a comparative evaluation and is thus an instance of evaluation. Consider something trivial as  $\mathbf{g_A}p_1$  = “advisor wishes the left door to be open” and  $\mathbf{g_A}p_2$  = “advisor wishes the right door to be open,” so that an evaluation  $\mathbf{e_A}p$  = “advisor prefers to open the left door instead of the right door” will indicate that  $\mathbf{g_A}p_1$  is more desirable than  $\mathbf{g_A}p_2$ .  $\mathbf{e_A}p$  thus states that there is a preference relation between  $\mathbf{g_A}p_1$  and  $\mathbf{g_A}p_2$ , indicating that the former is strictly more desirable than the latter.
- The **confidence** relation compares pairs of cues and/or advice in terms of uncertainty, to state if the events described in one of the two are more likely to occur than the events described in the other of the two. A confidence relation refers to a belief in events being more or less uncertain than others. The relation is thus stated via an instance of assumption. For  $\mathbf{a_A}p_1$  = “advisor believes that it will rain tomorrow” and  $\mathbf{a_A}p_2$  = “advisor believes that it will be sunny tomorrow,” so that the assumption  $\mathbf{a_A}p_3$  = “advisor believes that it is more likely that it will be raining tomorrow than that the sun will shine” gives a comparison of  $\mathbf{a_A}p_1$  and

<sup>2</sup> An inference step can be modeled in different ways, so that a more complex framework could include other inference rules, with or without modus ponens.

$\mathbf{a_A} p_2$  in terms of uncertainty about the events referred to in  $\mathbf{a_A} p_1$  and  $\mathbf{a_A} p_2$ , so that  $\mathbf{a_A} p_3$  states a confidence relation between  $\mathbf{a_A} p_1$  and  $\mathbf{a_A} p_2$ .

- Evaluation is not only comparative: individual cues/advice can be evaluated as well regardless of others. An instance of evaluation can thus indicate that a cue/advice is mandatory or optional. The **mandatory** relation is a unary relation, instance of evaluation which states that a cue/advice *must* be part of a solution to the advisor's problem in a given a-net. If the advisor cannot ignore the cue that the decision-maker desires to open the door, then that cue is mandatory: the advisor is thus obliged to engineer advice in such a way that the solution does not contradict this cue.
- In contrast to the mandatory relation, the **optional** relation indicates that it would be desirable for a solution to the advisor's problem to satisfy the optional cue/advice, but that the solution may be good enough, i.e., still acceptable if it contradicts the optional cue/advice. Perhaps while the advisor cannot ignore that the decision-maker desires to open the door, the advisor may if needed ignore that the decision-maker also desires that the window is open, so that in this case, the individual evaluation states that the latter cue is optional.

An a-net represents cues, advice, and relations between them. By including thereby the information that is relevant to the advisor in a given decision situation, an a-net includes all the information needed to define the advisor's problem for that decision situation. Analysis involves asking various questions on the basis of the a-net and searching answers thereto in a systematic way. The principal questions to ask are the following:

- *What is the advisor's problem in the given a-net?* While every a-net does define the advisor's problem, there is usually work to do on the a-net before it gives a clear statement of the advisor's problem, which once available, leads the advisor to search for candidate solutions, that is, to decide which advice to give. The cues that the advisor initially ascribes to the decision-maker need not be precise, complete, or detailed enough for the advisor to immediately formulate recommendations. Cues may need to be refined, and conflicts, preferences, mandatory and optional relations identified. Ascription is iterative, for the observed behavior of the decision-maker or the otherwise available information in the decision situation may require a revision of the initially ascribed cues. Apart from ascription, the advisor ought to determine – as in Example 6.1 – what his own goals, assumptions, evaluations, and commitments are, before he decides what advice to provide. The effects of these tasks are changes of the a-net, and thus changes to the statement of the advisor's problem.
- *What are the candidate solutions to the advisor's problem?* Given an a-net, the advisor's problem amounts to determine what recommendations to provide so as to satisfy the advisor's and his own preferred and mandatory goals, and not violate some of the assumptions. To search for candidate solutions is thus to study the a-net so as to identify advice to add to it to ensure some properties of the resulting a-net (e.g., that some goals are satisfied). Once the a-net includes conflicts, there will be more than one candidate solution, that is, more than

one alternative set of advice that could be given. Each of these candidate solutions will include different sets of cues. As preference, confidence, optional, and mandatory relations are defined over cues, it will be possible to compare candidate solutions in terms of desirability (via the preference, optional and mandatory relations), and in terms of confidence (via the corresponding relation).

- *Is a solution robust?* Roughly speaking, a solution is robust if it responds to its criticism. Establishing if a solution is robust requires its reformulation as a set of arguments. When a solution is understood as a set of arguments, its criticism can be reformulated as another set of arguments. It then follows that the latter arguments attack the former, conveying simply the relationship between arguments and counterarguments. Asking if a solution is robust thus amounts to determine whether the solution defends itself successfully against the information that counters it. If not, then it is possible to strengthen the solution, to change the advice that will be given so as to ensure that it responds to as many counterarguments as feasible in the given decision situation.
- *Is a solution clear?* Discussions in the fourth chapter pointed out that advice can be open to interpretation. To ask if a solution is clear is to study the openness of reference of a solution, and from the conclusions thus obtained, adjust the solution. This does not necessarily mean that the purpose of such an analysis is to eliminate vagueness, ambiguity, overgenerality, and so on. Rather, it may also involve reducing clarity if that seems to be the appropriate course of action to the advisor.

To make a framework that supports classification, modeling, and analysis, it is necessary to combine several facilities, as outlined earlier. One or more ontologies are needed to perform the classification of information. Modeling requires what is typically called a modeling language, usually defined as a mathematical formalism of some sort. Analysis demands that we are as clear as feasible on how to interpret, read models, and what to make of the information we find in, and can extract from them.

Most of the third and fourth chapters illustrated the challenges of conceptual analysis, of what engineered ontologies are, and what role they play in the analysis of advice. The construction and discussion of the cue and advice ontologies are themselves case studies in how to engineer ontologies. We shall thus stop here the discussions both of how to perform classification of the information relevant to the advisor, and of how to make ontologies that categorize this information – that was already done.<sup>3</sup>

---

<sup>3</sup> It is important to avoid the impression that ontology engineering amounts only to the definition of an ontology in the way discussed in Chaps. 3 and 4. What was presented there is one general way to do engineer ontologies. What else will be done depends on how the ontology is used. If it was necessary, for example, to record instances of the cue and advice ontologies in a computer, this would require that the ontologies be defined not only in natural language (as we did) but also in a language that is readable by a machine. Ontologies can also be written as mathematical theories so that their properties can be studied, such as checking if there are contradictions in the definitions of the concepts in the ontology. If interested in these topics, consult research

Once a classification scheme is available, the ensuing step is to precisely define what is precisely something that is called “a model” built out of this information, and what one can and cannot read from these models. A model is an artifact produced using a modeling language. The latter, just as a natural language such as English or French, must have some rules that need to be obeyed if those using it are to be successful in communication. Given that reference can be open, the question is not what one can read from a model, but how the modeling language should be defined to constrain the reading, interpretation of the models in ways that make them communicate as clearly as feasible what they are intended to communicate and limit misunderstanding. Getting closer to this aim requires that the modeling language is defined as a mathematical structure, so that models themselves are made by instantiating the concepts that make the language. A modeling language so defined differs from a natural language in that it does not tolerate loose use. Its components, what might amount to words and sentences of a natural language can be defined to be unambiguous in terms of how they relate to other elements of the modeling language. This does not mean that we can eliminate misunderstanding altogether, for it is still a person who reads the model and forms references herself, but it does mean that we can go one step further and define the modeling language in such a way that the reading of its components and models it produces are more constrained, and thus more precise in carrying their intended interpretation.<sup>4</sup>

### 6.3 Modeling Language

A modeling language is made of two parts, its syntax and its semantics. A language's *syntax* includes a collection of elementary signs (often called its *alphabet*) and rules (*grammar*) to state how signs can be combined into expressions. The purpose of *semantics* is to constrain the interpretation of the signs and expressions, by identifying the concepts to which the individual signs refer to, and by defining the reference relation between signs and concepts. There are consequently four parts to a modeling language:

$$\text{Modeling language} \stackrel{\text{def}}{=} \overbrace{\text{Alphabet} + \text{Grammar}}^{\text{Syntax}} + \overbrace{\text{Semantic domain} + \text{Semantic mapping}}^{\text{Semantics}}$$

Making a modeling language thus usually involves the definition of these four components. To model advice, we will be using the *Advice Modeling Language*

---

on the formalization of ontologies, logics for doing so, and applications in, e.g., facilitating the interoperability of databases, among others.

<sup>4</sup> Another effect is that such a language can be made readable to a machine, so that asking its models questions and finding answers can be automated, provided that the modeling language obeys some specific mathematical properties. Some are mentioned later in the text.

(AML), and we will introduce it below in three steps. This choice of presentation comes out of this section's double purpose. One aim is the definition of a conceptual toy – AML itself – with which to highlight some further ways of and considerations in any analysis of advice. The other benefit sought is to show how a simple modeling language can be defined from grounds up, so that the construction of AML acts as a case study in how to make a relatively simple modeling language. What is peculiar, though not uncommon about the presentation of AML below is that the language is constructed by detailing its syntax and semantics, i.e., by starting from very simple syntax and semantics which give a first modeling language that we will call AML2, then adding more details to it to make a second modeling language AML1, and then add further details to obtain AML.

Below, we start by looking into the syntax and semantics of AML2 which has only the conjunction, implication, inference, and conflict relations, and in which no distinction is made between kinds of propositions and expressions (i.e., no difference between a proposition being a goal or an assumption). AML1 will then add a classification of propositions and expressions on top of AML2. AML then finally adds the preference, confidence, mandatory, and optional relations to AML1.

### 6.3.1 AML2

The set of signs, i.e., the alphabet used in AML2 includes signs for propositions, for expressions, and for connectives.  $p, q, r, s$  and other lowercase letters of the Latin alphabet will each refer to a proposition, while lowercase letters of the greek alphabet, usually  $\phi, \psi, \gamma$ , will each refer to an expression. As we need potentially many of these, each sign for a proposition or an expression can be freely indexed (e.g.,  $p_5, \phi_{129}$ ).

Connectives are signs that wither refer to relations or are used to facilitate notation or reading.  $\wedge$  refers to the conjunction relation,  $\rightarrow$  to the implication relation. The horizontal bar separating the premises from the conclusion, i.e.,  $\frac{\text{premises}}{\text{conclusion}}$ , refers to the inference relation. There are only these three connectives and they cannot be indexed. The parentheses ( and ) are used to separate parts of expressions. The sign  $\perp$  is read “inconsistency” or equivalently “contradiction.”

**Definition 6.1. AML2 alphabet:** The alphabet of AML2 is exactly the following set of signs:

$$\begin{aligned} &\{p, p_1, \dots, p_n, q, q_1, \dots, q_n, \dots\} \\ &\cup \{\phi, \phi_1, \dots, \phi_n, \psi, \psi_1, \dots, \psi_n, \dots\} \\ &\cup \{\wedge, \rightarrow, (, ), \perp, \frac{\text{premises}}{\text{conclusion}}\} \end{aligned}$$

where  $n \geq 1$  is some positive natural number.



The next step is to state how the signs of the alphabet can be combined. The grammar tells us which such combinations are allowed in AML2.  $p \rightarrow q$  for example should be allowed, but  $p \wedge \rightarrow \wedge q$  cannot be.

**Definition 6.2. AML2 grammar:** The expressions of AML2 are those obtained by using the rules below, and only those, finitely many times:

- If  $p_1, \dots, p_n, q$  are signs that refer to propositions, then  $p_1 \wedge p_2 \wedge \dots \wedge p_n \rightarrow q$  is an expression.
- If  $p_1, \dots, p_n$  are signs that refer to propositions, then  $p_1 \wedge p_2 \wedge \dots \wedge p_n \rightarrow \perp$  is an expression.

Above,  $n \geq 1$  is some positive natural number.

An expression can also be called a *well-formed formula*, a name that emphasizes that the elements of the alphabet have been combined according to the rules of grammar, thereby giving something that is *well-formed*.<sup>5</sup>

Rules of grammar can be defined in another way, using what is called the Backus Naur form, in which Definition 6.2 is simply the following:

$$\phi ::= p_1 \wedge p_2 \wedge \dots \wedge p_n \rightarrow q \mid p_1 \wedge p_2 \wedge \dots \wedge p_n \rightarrow \perp \quad (6.1)$$

In (6.1), every indexed  $p$  and  $q$  refers to a proposition and  $\phi$  abbreviates any one of the two expressions on the right-hand side of  $::=$ .

*Remark 6.1.* Any indexed greek letter (e.g.,  $\gamma_{16}$ ) abbreviates an expression in AML2. This is a notational convention which will considerably simplify the discussions that follow.

The alphabet and grammar defined above form the syntax of AML2. The next step is to see what the semantic domain is for AML2.

It is convenient to think of the semantic domain as being a set of objects. It follows, given our discussions in Chap. 3, that the semantic domain must be made of concepts/universals and elementary particulars. Furthermore, an object will be either elementary, indivisible, or formed out of other elementary objects. What these elementary and complex objects are, i.e., how we will defined them depends essentially on what they should be to be relevant to the task at hand. Recall that the task in

---

<sup>5</sup> As a matter of notational convention, we have the following equivalences:

$$\begin{aligned} p_1 \wedge p_2 \wedge \dots \wedge p_n \rightarrow q &\equiv \bigwedge_{i=1}^n p_i \rightarrow q, \\ p_1 \wedge p_2 \wedge \dots \wedge p_n \rightarrow \perp &\equiv \bigwedge_{i=1}^n p_i \rightarrow \perp, \end{aligned}$$

where  $\equiv$  says that what is written on its left-hand side is always interchangeable (i.e., is the same) with what is written on its right-hand side, and vice versa.

question is the definition of a modeling language which would represent information useful in the definition and resolution of the advisor's problem.

What we are interested to find in the semantic domain are *candidate solutions*. The semantic domain should thus be made out of candidate solutions, which are instances of a concept of candidate solution, and it is these that we need to define. Several questions ensue. Is a candidate solution elementary or made out of some other particulars which are elementary? Are there different kinds of candidate solutions, so that the candidate solution concept can be specialized? It is practical to take the other way around: instead of asking these questions right at the outset, we can start by defining elementary particulars and universals in the semantic domain by considering what we already introduced, namely the alphabet and grammar of AML2, then see how to make the candidate solution concept out of the elementary universals and particulars in the semantic domain.

The alphabet has signs for propositions, expressions, and connectives. Every sign for a proposition refers to a particular proposition, that is, something which is the object of a psychological attitude. A first set of elementary constituents of the semantic domain are then propositions, which for all practical purposes can be viewed as arbitrary sentences of a natural language, such as English. The sentence "Door is open" is thus a particular proposition, an instance of the concept of proposition. We can use the sign/concept/object triangle to illustrate the reference and instantiation relationships between signs for propositions in the alphabet of AML2 and their targets in the semantic domain, the particular propositions and the concept of proposition. Figure 6.4 illustrates this, showing at the same time what participates in the syntax and what in the semantics of AML2.

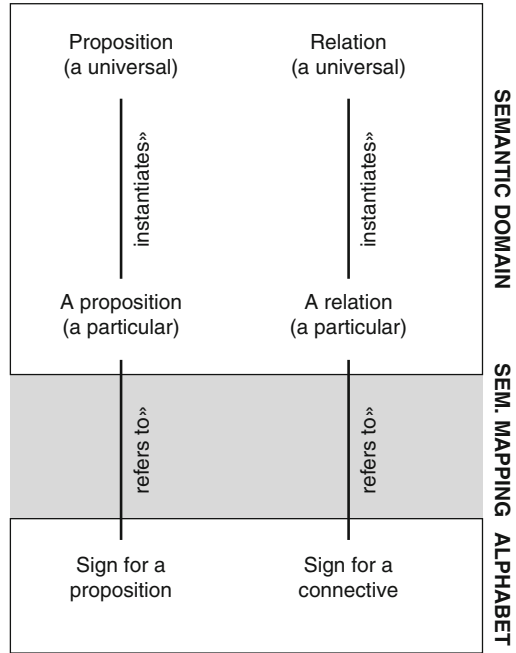
The reference relation acts as the bridge between the alphabet and the semantic domain, playing thereby the role of the semantic mapping component of AML2. In addition to propositions, we have four elementary relations in the semantic domain. The counterpart of a relation in the alphabet is a connective sign, of which we have four in AML2:  $\wedge$ ,  $\rightarrow$ ,  $\mathbf{I}$  and  $\mathbf{C}$ . We consequently have instances of the conjunction, implication, inference, and conflict relations in the semantic domain, relations which stand between particular propositions. Figure 6.4 consequently shows the part of the semantic domain which includes only elementary objects, be they universals or particulars. What that figure does not show are the objects in the semantic domain which are made by combining the elementary objects, the propositions, and relations.

We take the conjunction and implication relations as primitives, and leave them undefined, so that we ought to define the inference and conflict relations.

**Definition 6.3. Inference relation ( $\mathbf{I}$ ).** If the semantic domain includes:

1. the propositions referred to by  $p_1, \dots, p_n$ ,
2. the conjunction relation between propositions  $p_1, \dots, p_n$ ,
3. the implication relation according to which  $\bigwedge_{i=1}^n p_i$  implies  $q$

**Fig. 6.4** Reference and instantiation relations between signs in the alphabet of a modeling language and the elementary objects in the semantic domain of that language



then the semantic domain also includes the proposition referred to by  $q$  and it includes the inference relation

$$\frac{p_1 \quad \dots \quad p_n \quad \bigwedge_{i=1}^n p_i \rightarrow q}{q}$$

in which the propositions  $p_1, \dots, p_n$  and the expression  $\bigwedge_{i=1}^n p_i \rightarrow q$  are premises and  $q$  is the conclusion of the inference relation.

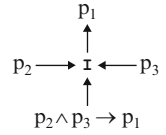
*Example 6.2.* Niccolò Machiavelli's *The Prince* is a book of advice on how to acquire and hold political power. In his third chapter, he argues how a prince is to keep the principalities he inherited, and suggests:

It is my opinion that it is less difficult to preserve a state which is hereditary and accustomed to the family of their prince than one which is recently established: it is enough not to neglect the constitutional arrangements made by one's predecessors, and then adapt one's conduct to circumstances as they arise; (Machiavelli 2009, Sect. 2)

We can extract the following propositions and refer them to signs for propositions:

- $p_1$  abbreviates “to preserve a state which is hereditary and accustomed to the family of their prince”.
- $p_2$ : “it is enough not to neglect the constitutional arrangements made by one's predecessors”.
- $p_3$ : “adapt one's conduct to circumstances as they arise.”

**Fig. 6.5** Advisor's reasoning in Example 6.2 shown as a graph



The entire passage that we quoted above can then be rewritten as  $p_2 \wedge p_3 \rightarrow p_1$ . According then, to Definition 6.3 if the semantic domain has referents of  $p_2$  and  $p_3$ , and of  $p_2 \wedge p_3 \rightarrow p_1$ , then the semantic domain will also have the inference relation, in which the premises are the set  $\{p_2, p_3, p_2 \wedge p_3 \rightarrow p_1\}$ , while the conclusion is  $p_1$ , and will include that conclusion. In other words, what we simply established is that the passage leads us to conclude that if the prince does not neglect the constitutional arrangements made by one's predecessors (i.e.,  $p_2$ ) and he adapts his conduct to circumstances as they arise ( $p_3$ ), then the prince will preserve a state which is hereditary and accustomed to the family of this prince (i.e.,  $p_1$ ). Figure 6.5 shows the propositions and the inference relation between them.

The idea of the inference relation is that it will be present in the semantic domain and between some premises and a conclusion whenever specific conditions are satisfied, which are listed in Definition 6.3. The first condition gives some of the necessary premises, while the second condition indicates that there must already be a conjunction relation between these premises. The third condition further requires that there is an implication relation in the semantic domain. When the three conditions are satisfied, the semantic domain is said to contain also an inference relation between the given premises and conclusion.

The conflict relation shares some of the conditions with the inference relation, so that their definitions are similar. The conclusion in the case of the conflict relation is inconsistency. The relation thus stands between propositions which together are inconsistent, or equivalently, are contradictory.

**Definition 6.4. Conflict relation (C).** If the semantic domain includes:

1. the propositions referred to by  $p_1, \dots, p_n$ ,
2. the conjunction relation between propositions  $p_1, \dots, p_n$ ,
3. the implication relation according to which  $\bigwedge_{i=1}^n p_i$  implies inconsistency ( $\perp$ )

then the semantic domain also includes the conflict relation

$$\frac{p_1 \quad \dots \quad p_n \quad \bigwedge_{i=1}^n p_i \rightarrow \perp}{\perp}$$

in which the propositions  $p_1, \dots, p_n$  and the expression  $\bigwedge_{i=1}^n p_i \rightarrow \perp$  are premises and  $\perp$  is the conclusion of the conflict relation.

*Example 6.3.* A mixed principality, Machiavelli says, is made by combining principalities previously held by the prince with those newly acquired. He advises the following:

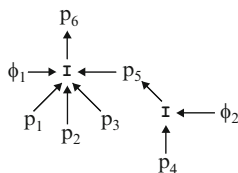
[A]nyone who is in a province with an alien culture ought to make himself head and defender of lesser rulers who are round about, and take pains to weaken those who are more powerful, and take care that no unforeseen event may lead to the appearance of a foreigner as powerful as he is. [...] And the way of things is such that, as soon as a foreign potentate enters a province, all those there who are less powerful come under his influence, moved by the envy which they bear to whoever has been in power over them; and the result is that he needs to make no effort to win these less powerful people, since they are straight away happy to become one with the state he has conquered. He need only take care that they do not acquire too much power or authority. (Machiavelli 2009, Sect. 3)

We can pick out the following propositions in the passage above:

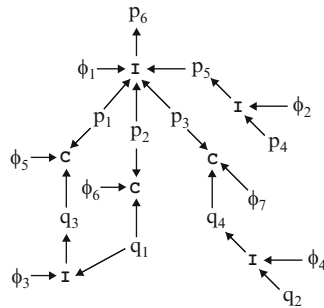
- $p_1$ : one ought to make himself head and defender of lesser rulers who are round about.
- $p_2$ : one ought to weaken those who are more powerful.
- $p_3$ : one ought to take care that no unforeseen event may lead to the appearance of a foreigner as powerful as he is.
- $p_4$ : as soon as a foreign potentate enters a province, all those there who are less powerful come under his influence, moved by the envy which they bear to whoever has been in power over them.
- $p_5$ : one needs to make no effort to win these less powerful people, since they are straightaway happy to become one with the state he has conquered.
- The following is implicit, as it is the purpose of doing as suggested above:  $p_6$  the prince will keep the mixed principality without much difficulties and troubles.

We can see from the passage that the following are also seem to be assumed by Machiavelli: (1) if  $p_4$  then  $p_5$ , i.e.,  $p_4 \rightarrow p_5$ , as well as (2) if  $p_1$  and  $p_2$  and  $p_3$  and  $p_5$  then  $p_6$ . The reasoning that the passage displays is shown in the graph in Fig. 6.6.

Another option a prince might follow would be to overthrow a lesser ruler who is round about ( $q_1$ ) and cooperate with a mighty foreign ruler to overthrow other lesser rulers ( $q_2$ ). Machiavelli further argues that if the prince overthrows a lesser ruler, then he will lose the trust of other lesser rulers in his mixed principality ( $q_3$ ), i.e.,  $q_1 \rightarrow q_3$ . Moreover, if the prince cooperates with a mighty foreign ruler to overthrow lesser ones, then he will introduce into his mixed principality that mighty foreign ruler ( $q_4$ ), i.e.,  $q_2 \rightarrow q_4$ .



**Fig. 6.6** Advisor's reasoning in Example 6.3 shown as a graph, before the contradicting information is introduced. In the figure:  $\phi_1 \equiv p_1 \wedge p_2 \wedge p_3 \wedge p_5 \rightarrow p_6$  and  $\phi_2 \equiv p_4 \rightarrow p_5$



**Fig. 6.7** Advisor's reasoning in Example 6.3 shown as a graph, after the contradicting information is introduced. In the figure:  $\phi_1 \equiv p_1 \wedge p_2 \wedge p_3 \wedge p_5 \rightarrow p_6$ ,  $\phi_2 \equiv p_4 \rightarrow p_5$ ,  $\phi_3 \equiv q_1 \rightarrow q_3$ ,  $\phi_4 \equiv q_2 \rightarrow q_4$ ,  $\phi_5 \equiv p_1 \wedge q_3 \rightarrow \perp$ ,  $\phi_6 \equiv p_2 \wedge q_1 \rightarrow \perp$ , and  $\phi_7 \equiv p_3 \wedge q_4 \rightarrow \perp$

It is quite clear that (1)  $p_1$  and  $q_3$  are contradictory, which we write  $p_1 \wedge q_3 \rightarrow \perp$ , (2)  $p_2$  and  $q_1$  are contradictory, i.e.,  $p_2 \wedge q_1 \rightarrow \perp$ , and (3)  $p_3$  and  $q_4$  are contradictory, i.e.,  $p_3 \wedge q_4 \rightarrow \perp$ . We thus have three conflicts, as shown in Fig. 6.7.

We now have an alphabet, a grammar, and the part of the semantic domain which includes elementary objects. The ensuing question is how we can combine these components to build more complex structures, and how they fit into the idea of the semantic domain as of a set of “candidate solutions.”

Example 6.3 was a case of a ruler who was presented with two ways of managing newly acquired lands. Contradictions ensure that if he follows  $q_1$  and  $q_2$ , then he cannot at the same follow do  $p_1$ – $p_4$ . Even if he is advised both, he cannot do both, or at least that is what the contradictions say. It then appears reasonable to postulate that *a candidate solution cannot include contradictions*. This does not mean that all advice must be internally logically consistent, i.e., contain no contradictions. Rather, it means that when contradictory advice is given to the decision-maker, then actually several sets of advice, each internally consistent, are being given. Each internally consistent advice will consequently recommend one alternative way of acting, thinking, or otherwise, to the decision-maker.

If every candidate solution ought to be free of conflicts, then we can build a candidate solution by putting together the information that includes no contradictions, and which does not allow us to conclude contradictions.

Before we do start making internally consistent sets of information, it is important to make a distinction at this point, *between the information that is given and the information that is inferred from the information that is given*. To see this difference, consider Example 6.2 again: we started there from a passage from Machiavelli, and rewrote it using AML2 to highlight the conditional and inference relations in his recommendation. The information that he *gives* are  $p_2$ ,  $p_3$ , and  $p_2 \wedge p_3 \rightarrow p_1$ . What he concludes on the basis of the give information is  $p_1$ . According to that excerpt from *The Prince*, if the prince does not neglect the constitutional arrangements made by one's predecessors ( $p_2$ ) and adapts his conduct to circumstances as

they arise ( $p_3$ ), then he will preserve a state which is hereditary and accustomed to the prince's family ( $p_1$ ). If we avoid debating whether Machiavelli was being sarcastic, we can say that it  $p_2$ ,  $p_3$ , and  $p_2 \wedge p_3 \rightarrow p_1$  are enough for him to conclude  $p_1$ . In other words, he started from some information and drew conclusions, while we represented the reasoning he thereby seems to have done by a single application of modus ponens:

$$\frac{p_2 \quad p_3 \quad p_2 \wedge p_3 \rightarrow p_1}{p_1}.$$

If a decision-maker follows exactly what Machiavelli is saying, then that decision-maker will himself conclude  $p_1$  once he is given  $p_2$  and  $p_3$  and if he, as Machiavelli, takes for granted that  $p_2 \wedge p_3 \rightarrow p_1$ . The importance of the distinction between given and inferred information becomes clearer when several steps of inference are involved, as in Example 6.3 with conflicting recommendations and Example 6.1 with the door to open.

Suppose that  $\Gamma$  is some arbitrary set of propositions and expressions; we say arbitrary as it is unimportant what is actually in it for our discussion. Let  $p$  be some proposition; whether it is in  $\Gamma$  or not in  $\Gamma$  is also unimportant (it may be, but need not). If  $p$  can be inferred from the propositions and expressions in  $\Gamma$ , then the two stand in the relation called the *consequence relation*. The sign used to refer to it is  $\vdash$ , so we write  $\Gamma \vdash p$  to say that  $p$  can be inferred from (equivalently, is a consequence of)  $\Gamma$ .

**Definition 6.5. Consequence relation ( $\vdash$ ):** Let  $\Gamma$  be some nonempty set of propositions and/or expressions in AML2. Let  $x$  be an abbreviation of either a proposition  $p$  or of inconsistency, i.e.,  $x \in \{p, \perp\}$ :

1.  $\Gamma \vdash p$  if  $p \in \Gamma$ , or
2.  $\Gamma \vdash x$  if for all  $i$  such that  $1 \leq i \leq n$ ,  $\Gamma \vdash q_i$  and  $\bigwedge_{i=1}^n q_i \rightarrow x$  is in  $\Gamma$ , i.e.,  $(\bigwedge_{i=1}^n q_i \rightarrow x) \in \Gamma$ .

Definition 6.5 takes a set of propositions and/or expressions  $\Gamma$  and gives conditions which need to be satisfied if we are to write that either the proposition  $p$  or inconsistency are consequences of  $\Gamma$ . The first condition says that if  $p$  is in  $\Gamma$ , then  $\Gamma \vdash p$ . The second condition is satisfied if we can apply modus ponens some number of times over  $\Gamma$  and conclude  $x$  (where  $x$  is either  $p$  or  $\perp$ ).

*Example 6.4.* Let  $\Gamma_1 = \{p, p \rightarrow q\}$ . What are all the consequences of  $\Gamma_1$ ? According to the first condition in Definition 6.5,  $p$  is a consequence of  $\Gamma_1$ , i.e.,  $\Gamma_1 \vdash p$ . Since  $\Gamma_1 \vdash p$  and  $p \rightarrow q$  is in  $\Gamma_1$ , the second condition in Definition 6.5 tells us that  $q$  is also a consequence of  $\Gamma_1$ , i.e.,  $\Gamma_1 \vdash q$ . The set of all consequences of  $\Gamma_1$  is, then,  $\{p, q\}$ .

Suppose that  $\Gamma_2 = \{p, p \rightarrow q, q \rightarrow r, r \rightarrow s\}$ . What are the consequences of  $\Gamma_2$ ? Clearly,  $p$  is, since it is in  $\Gamma_2$ . By applying the second condition from Definition 6.5, we see that  $q$  is also a consequence of  $\Gamma_2$ . Since  $\Gamma_2 \vdash q$ , it follows, again by the second condition, that  $r$  is also a consequence of  $\Gamma_2$ . By the same rationale,  $s$  is a consequence of  $\Gamma_2$ . The set of all consequences of  $\Gamma_2$  is  $\{p, q, r, s\}$ .

Let  $\Gamma_3 = \{p, p \wedge q \rightarrow r\}$ . Remark that here,  $r$  is not a consequence of  $\Gamma_3$  as  $q$  is not a given. We can thus write that  $\Gamma_3 \not\vdash r$ .

The set of all consequence of some set of propositions and/or expressions  $\Gamma$  is called *the closure of  $\Gamma$* .<sup>6</sup>

**Definition 6.6. Closure:** Let  $\Gamma$  be a set of propositions and/or expressions in AML2. The closure of  $\Gamma$ , written  $\text{Cl}(\Gamma)$  is the set of all consequences of  $\Gamma$ , i.e.,  $\text{Cl}(\Gamma) = \{p \mid \Gamma \vdash p\}$ .

Before we move further closer to the candidate solution concept, some technical remarks are in order. Perhaps, the closest well-known logic that AML2 may seem to have similarities with is classical propositional logic (hereafter CPL). Differences between the two are significant, however, both at the level of syntax and of the semantics. Propositions and expressions can be negated in CPL, while they cannot in AML2. CPL has more inference rules, while AML2 has only modus ponens. It follows that some forms of inference cannot be represented in AML2, while they can in CPL.<sup>7</sup> A beneficial trait of AML2 compared to CPL is that  $\vdash$  is paraconsistent, while the consequence relation in CPL is not.

To be *paraconsistent* (Priest and Tanaka 2009), a consequence relation must not allow any conclusion to be drawn from contradiction. Suppose that  $\Gamma = \{p, q, p \wedge q \rightarrow \perp\}$ , so that  $p$  and  $q$  are contradictory. If  $\vdash$  were *not* paraconsistent, then we could conclude anything from  $\Gamma$ : we could conclude some  $r$ , i.e., we could write  $\Gamma \vdash r$ , and this regardless of what is stated in the proposition that  $r$  refers to (e.g.,  $r$  could abbreviate “the Earth is flat”). A consequence relation which is *not paraconsistent* is also called an *explosive* consequence relation, because anything is a consequence of  $\Gamma$  for an explosive consequence relation. A paraconsistent consequence relation is not explosive: it is said to reject the principle that anything follows from a contradiction (i.e., it rejects the *ex contradictione quodlibet* principle). That  $\vdash$  is paraconsistent can be shown in a straightforward way, as the proof of Proposition 6.1 illustrates.

**Proposition 6.1.** *The consequence relation  $\vdash$  is paraconsistent.*

*Proof.* The proof is by contradiction.  $\vdash$  is paraconsistent if and only if for any  $p, q, r$ , such that  $p \neq q \neq r$ ,  $r$  is not the consequence of  $\Gamma$ , where  $\Gamma = \{p, q, p \wedge q \rightarrow \perp\}$ : i.e.,  $\vdash$  is paraconsistent if and only if  $\Gamma \not\vdash r$ . Suppose the opposite: that  $\vdash$  is not paraconsistent. If so, then for any  $p, q, r$ , such that  $p \neq q \neq r$ , we have that  $\Gamma \vdash r$ , which contradicts the first condition in Definition 6.5, since  $r \notin \Gamma$ . It also contradicts the second condition in Definition 6.5, because there is no expression in  $\Gamma$  which is of the form  $\dots \rightarrow r$ . Since saying that  $\vdash$  is not paraconsistent contradicts Definition 6.5, we conclude that  $\vdash$  is paraconsistent.

<sup>6</sup> In the definition, we write  $\{p \mid \Gamma \vdash p\}$ , where  $\mid$  to say “the set of all propositions which are consequences of  $\Gamma$ .”

<sup>7</sup> Also note that  $\vdash$  is sound with regards to standard entailment in CPL, but is incomplete in two ways: it only considers deducing positive propositions (not negated ones), and no ordinary proofs based on arguing by contradiction go through.



We now have the syntax and the elementary parts of semantics needed to define the candidate solution concept. Recall above that we want no contradictions in a candidate solution. We will consequently build a candidate solution out of small consistent bundles of information, which we call *picosolutions*.

*Remark 6.2.* We denote all the given information, i.e., the given (and not inferred) propositions and expressions by  $\mathbf{S}(\mathcal{A})$ . Remark from Definition 6.5 that we cannot have an expression on the right-hand side of  $\vdash$ , i.e., we cannot conclude expressions. Every expression is consequently a given, and we write  $\mathbf{E}(\mathcal{A})$  to refer to the set of all expressions. As all expressions are given, it is clear that  $\mathbf{E}(\mathcal{A}) \subseteq \mathbf{S}(\mathcal{A})$ .

**Definition 6.7. Picosolution for a proposition.** Let  $\mathbf{S}(\mathcal{A})$  be the set of all given propositions and expressions, and  $\mathbf{E}(\mathcal{A})$  the set of all given expressions. A set  $\Gamma$  of propositions and expressions is a picosolution for a proposition  $p$ , written  $\langle \Gamma, p \rangle$ , if and only if:

1.  $p$  is a consequence of  $\Gamma$ :  $\Gamma \vdash p$ ,
2. every proposition and expression in  $\Gamma$  is a given piece of information:  $\Gamma \subseteq \mathbf{S}(\mathcal{A})$ ,
3. no contradiction is a consequence of  $\Gamma$ :  $\Gamma \not\vdash \perp$ ,
4. all expressions are in  $\Gamma$ :  $\mathbf{E}(\mathcal{A}) \subseteq \Gamma$ ,
5. there are no unnecessary propositions in  $\Gamma$ : there is no  $\Gamma_i \subset \Gamma$  such that both  $\Gamma_i \vdash p$  and  $\mathbf{E}(\mathcal{A}) \subseteq \Gamma_i$ .

What a picosolution for some  $p$  does is that it puts together a consistent set of given information from which  $p$  can be concluded. A picosolution for  $p$  thus includes the information that supports  $p$ . When there are different ways to conclude  $p$ , there is more than one picosolution for  $p$ : e.g., if  $\mathbf{S}(\mathcal{A}) = \{p, p \rightarrow q, r, r \rightarrow q\}$ , then there are two picosolutions for  $q$ , one being  $\{\{p, p \rightarrow q, r \rightarrow q\}, q\}$  and the other  $\{\{p \rightarrow q, r, r \rightarrow q\}, q\}$ .

The first condition in Definition 6.7 makes sure that a picosolution includes the information that supports  $p$ , i.e., the information from which we can conclude  $p$ . According to the second condition, we want only the given information in the  $\Gamma$  of  $\langle \Gamma, p \rangle$ . Since we wish to exclude contradictions from a picosolution, we must include the third condition. If the fourth condition was missing, it would be possible to make picosolutions by simply ignoring expressions that state conflicts, those of the form  $\bigwedge_{i=1}^n p_i \rightarrow \perp$ . The last condition in Definition 6.7 says that  $\Gamma$  in  $\langle \Gamma, p \rangle$  includes on top of all expressions  $\mathbf{E}(\mathcal{A})$  only those propositions which are necessary and sufficient to conclude  $p$ . The last condition specifically says that there is no set  $\Gamma_i$  of information which is smaller than  $\Gamma$  and from which we can still conclude  $p$  – if such a set existed, it would mean that  $\Gamma$  includes propositions which are unnecessary to our reaching the conclusion  $p$ , and there is no need to have these additional propositions in the picosolution.

*Example 6.5.* Consider again Example 6.2 (see Fig. 6.5 as well). We can say the following about the picosolutions there:

- $\mathbf{E}(\mathcal{A}) = \{\phi_1, \phi_2\} = \{p_1 \wedge p_2 \wedge p_3 \wedge p_5 \rightarrow p_6, p_4 \rightarrow p_5\}$
- $\mathbf{S}(\mathcal{A}) = \mathbf{E}(\mathcal{A}) \cup \{p_1, p_2, p_3, p_4\}$

- $\langle E(\mathcal{A}) \cup \{p_4\}, p_5 \rangle$
- $\langle E(\mathcal{A}) \cup \{p_1, p_2, p_3, p_4\}, p_6 \rangle$

A picosolution is a building block of candidate solutions. Since we want no contradictions in a candidate solution, we put together picosolutions which are consistent. The definition of the candidate solution concept is consequently straightforward.

**Definition 6.8. Candidate solution:** A set  $\mathcal{S}$  of picosolutions is a candidate solution if and only if  $\mathcal{S} = \{\langle \Gamma_i, p_i \rangle \mid \bigcup_{i=1}^n \Gamma_i \not\vdash \perp\}$ .

*Example 6.6.* Consider again Example 6.3 and in particular Fig. 6.7 in which there are conflicts. What are the solutions in that Example? The following are some of the solutions:

- In that Example  $E(\mathcal{A}) = \{\phi_1, \dots, \phi_7\}$  and  $S(\mathcal{A}) = E(\mathcal{A}) \cup \{p_1, p_2, p_3, p_4, q_1, q_2\}$ . The smallest solutions each contain one picosolution only, so that  $\langle E(\mathcal{A}) \cup \{p_1\}, p_1 \rangle$  is a candidate solution, and so is each of the following:  $\langle E(\mathcal{A}) \cup \{p_2\}, p_2 \rangle$ ,  $\langle E(\mathcal{A}) \cup \{p_3\}, p_3 \rangle$ ,  $\langle E(\mathcal{A}) \cup \{p_4\}, p_4 \rangle$ ,  $\langle E(\mathcal{A}) \cup \{q_1\}, q_1 \rangle$ , and  $\langle E(\mathcal{A}) \cup \{q_2\}, q_2 \rangle$ .
- $\langle E(\mathcal{A}) \cup \{p_4\}, p_5 \rangle$  alone is a candidate solution.
- $\langle E(\mathcal{A}) \cup \{p_1, p_2, p_3, p_4\}, p_6 \rangle$  alone is a candidate solution.
- $\langle E(\mathcal{A}) \cup \{q_1\}, q_3 \rangle$  alone is a candidate solution.
- $\langle E(\mathcal{A}) \cup \{q_2\}, q_4 \rangle$  alone is a candidate solution.
- $\{\langle E(\mathcal{A}) \cup \{p_4\}, p_5 \rangle, \langle E(\mathcal{A}) \cup \{p_1, p_2, p_3, p_4\}, p_6 \rangle\}$  is a candidate solution which includes two picosolutions.
- $\{\langle E(\mathcal{A}) \cup \{q_1\}, q_3 \rangle, \langle E(\mathcal{A}) \cup \{q_2\}, q_4 \rangle\}$  is a candidate solution which includes two picosolutions.
- $\{\langle E(\mathcal{A}) \cup \{p_4\}, p_5 \rangle, \langle E(\mathcal{A}) \cup \{q_1\}, q_3 \rangle\}$  is a candidate solution.
- $\{\langle E(\mathcal{A}) \cup \{p_4\}, p_5 \rangle, \langle E(\mathcal{A}) \cup \{q_2\}, q_4 \rangle\}$  is a candidate solution.
- $\{\langle E(\mathcal{A}) \cup \{p_4\}, p_5 \rangle, \langle E(\mathcal{A}) \cup \{q_1\}, q_3 \rangle, \langle E(\mathcal{A}) \cup \{q_2\}, q_4 \rangle\}$  is a candidate solution.

Following Definition 6.8 we cannot have contradictions in a candidate solution, so that, e.g., the set  $\{\langle E(\mathcal{A}) \cup \{p_1, p_2, p_3, p_4\}, p_6 \rangle, \langle E(\mathcal{A}) \cup \{q_1\}, q_3 \rangle\}$  is not a candidate solution.

There can be many solutions for a given set of propositions and expressions. Example 6.6 is a clear illustration of this, and is especially striking as the given set of information is very small.

We cannot in AML2 compare solutions in terms of preference, confidence or otherwise: AML2 does not include the preference and other relations relevant for such a comparison. We can, however, provide a criterion which is independent of these relations, and which ranks solutions in a way that lets us identify those that are – regardless of individuals' preferences or confidence – more significant than others. The idea is simple: *because some solutions are subsets of others, we can use the set inclusion relation ( $\subseteq$ ) relation for the comparison of solutions.* If we

compare thus two solutions, and one is the subset of the other, it seems more relevant to adopt the larger one. Doing otherwise means that we choose to ignore some of the information that is consistent with the chosen candidate solution.

*Example 6.7.* Example 6.6 gave a number of solutions, among them  $\mathcal{S}_1 = \{\langle E(\mathcal{A}) \cup \{q_1\}, q_3 \rangle, \langle E(\mathcal{A}) \cup \{q_2\}, q_4 \rangle\}$  and  $\mathcal{S}_2 = \{\langle E(\mathcal{A}) \cup \{q_2\}, q_4 \rangle\}$  and  $\mathcal{S}_3 = \{\langle E(\mathcal{A}) \cup \{q_1\}, q_3 \rangle\}$ . Since  $\mathcal{S}_2 \subset \mathcal{S}_1$  and  $\mathcal{S}_3 \subset \mathcal{S}_1$ , the largest among these three solutions and the most interesting one is  $\mathcal{S}_1$ .

There is then the candidate solution

$$\mathcal{S}_4 = \{\langle E(\mathcal{A}) \cup \{p_4\}, p_5 \rangle, \langle E(\mathcal{A}) \cup \{p_1, p_2, p_3, p_4\}, p_6 \rangle\},$$

as well as, e.g.,  $\mathcal{S}_5 = \langle E(\mathcal{A}) \cup \{p_4\}, p_5 \rangle$ , and  $\mathcal{S}_6 = \langle E(\mathcal{A}) \cup \{p_1, p_2, p_3, p_4\}, p_6 \rangle$ .  $\mathcal{S}_4$  is the biggest among the three, as  $\mathcal{S}_5 \subset \mathcal{S}_6 \subset \mathcal{S}_4$ .

We specialize the candidate solution concept to capture the idea above, that we are interested more in so to speak bigger solutions and not the smaller ones.

**Definition 6.9. Maximal candidate solution:** A candidate solution  $\mathcal{S}$  is maximal if and only if there is no other candidate solution  $\mathcal{S}_i$  such that  $\mathcal{S} \subset \mathcal{S}_i$ .

Whether a candidate solution is maximal depends on the presence of its supersets. It follows – and Example 6.7 illustrates this – that there can be more than one candidate solution which we can call maximal. To pick out a candidate solution, we cannot rely *only* on the maximality criterion, i.e., to seek only maximal solutions. Additional relations are needed, hence the preference, mandatory, optional and confidence relations in AML.

### 6.3.2 AML1

The first variant of the advice modeling language AML2 does not make the distinction between the mode and content in communication. We cannot say in AML2 that a proposition is a goal, and that another one is an assumption. What AML1 does is that it takes AML2 and sorts the propositions and expressions, to distinguish those which are cues from those which are advice, then distinguish among specializations of the cue and advice concepts.

To say which propositions and expressions are cue and which are advice, we introduce classification rules. We have one classification rule per concept from the cue and advice ontologies. The rule simply checks whether a proposition or expression satisfies the identity criteria of a concept from the two ontologies. The job was already done: the definition of each concept from these two ontologies already gives the identity criteria of interest. The rules take identity criteria from Definitions 5.1–5.12.

Chapter 5 defined the cue and advice ontologies. Cues were defined as the advisor's ascriptions of intentional states to the decision maker. Advice were the information that the advisor communicates to support or attack these cues, thereby aiming to either reinforce or challenge them.

To have the classification of propositions and expressions in AML1 new signs are needed. Each acts as a label to a proposition or expression. As there are assumptions, goals, evaluations, and commitments, at least four signs are needed. Advice was defined as a triple  $(A, x, D)$ , where **A** stood for the advisor,  $x$  for the advisor's speech act, and **D** for the decision maker who is the target of the speech act. Advice was then  $(A, x, c, D)$ , where  $c$  was the cue which  $x$  either supports or attacks. Four signs cannot be enough then, as the  $x$  can be seen as merely a manifestation in communication of the advisor's own intentional states. We can have eight signs then, four for cues that **A** holds about **D**, and four signs for the advisor's own assumptions, goals, evaluations and commitments. We will stay in the simpler case where we have one advisor and one decision-maker, so that the signs are not indexed.

**Definition 6.10. AML1 alphabet:** The alphabet of AML1 is exactly the following set of signs:

$$\{p, p_1, \dots, p_n, q, q_1, \dots, q_n, \dots\} \cup \{\phi, \phi_1, \dots, \phi_n, \psi, \psi_1, \dots, \psi_n, \dots\} \\ \cup \{\mathbf{a}_D, \mathbf{g}_D, \mathbf{e}_D, \mathbf{c}_D, \mathbf{a}_A, \mathbf{g}_A, \mathbf{e}_A, \mathbf{c}_A\} \cup \{\wedge, \rightarrow, (, ), \perp, \frac{\text{premises}}{\text{conclusion}}\}$$

where  $n \geq 1$  is some positive natural number.

Compared to the alphabet of AML2, we simply added eight signs. The grammar of AML2 says how these signs can be combined with propositions and expressions.

**Definition 6.11. AML1 grammar:** The labeled propositions and expressions of AML1 are those obtained by using the rules below, and only those, finitely many times:

- If  $p$  is sign that refers to a proposition, then  $\mathbf{a}_A p, \mathbf{g}_A p, \mathbf{e}_A p, \mathbf{c}_A p, \mathbf{a}_D p, \mathbf{g}_D p, \mathbf{e}_D p, \mathbf{c}_D p$  are labeled propositions.
- If  $pl_1, \dots, pl_n, pl$  are signs that refer to labeled propositions, then  $pl_1 \wedge pl_2 \wedge \dots \wedge pl_n \rightarrow pl$  is an expression.
- If  $pl_1, \dots, pl_n$  are signs that refer to labeled propositions, then  $pl_1 \wedge pl_2 \wedge \dots \wedge pl_n \rightarrow \perp$  is an expression.
- If  $\phi$  is an expression, then  $\mathbf{a}_D \phi$  and  $\mathbf{a}_A \phi$  are labeled expressions.

Above,  $n \geq 1$  is some positive natural number.

Instead of dealing with propositions and expressions as in AML2, the interest in AML1 is to deal with *labeled* propositions and expressions. While any proposition can be labeled with any of the eight labels, the same does not apply to expressions. Every expression in AML1 is always an instance of the assumption concept.

Definition 6.11 in Backus Naur form is as follows:

$$pl ::= \mathbf{a_A}p \mid \mathbf{g_A}p \mid \mathbf{e_A}p \mid \mathbf{c_A}p \mid \mathbf{a_D}p \mid \mathbf{g_D}p \mid \mathbf{e_D}p \mid \mathbf{c_D}p, \quad (6.2)$$

$$\phi ::= \bigwedge_{i=1}^n pl_i \rightarrow pl \mid \bigwedge_{i=1}^n pl_i \rightarrow \perp, \quad (6.3)$$

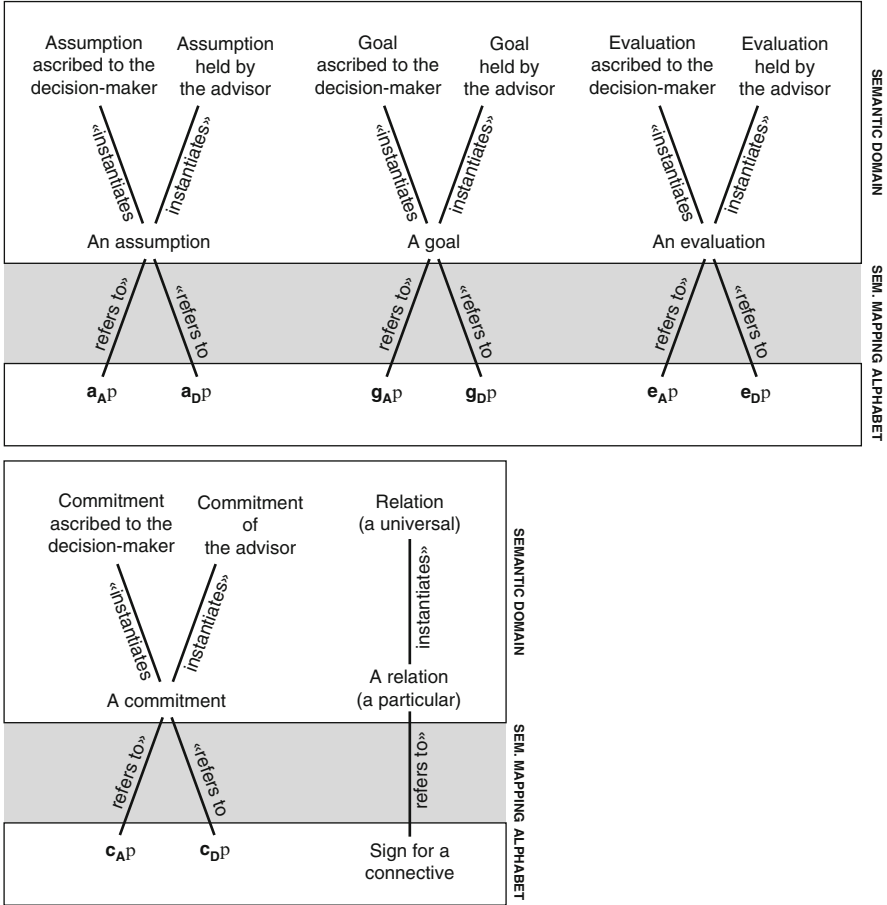
$$el ::= \mathbf{a_D}\phi \mid \mathbf{a_A}\phi. \quad (6.4)$$

The next step follows the path we took in defining the semantic domain and semantic mapping for AML2. We still have a semantic domain in which there are propositions and relations between propositions to form expressions, but there is also a classification of propositions and of expressions. We consequently partition the set of all propositions according to the cue and advice ontologies.

Figure 6.8 illustrates the alphabet, semantic mapping, and the semantic domain in AML1. Signs for connectives refer to relations, and AML1 includes the same relations as AML2, namely the conjunction, implication, inference, and conflict relations. Where AML2 and AML1 differ is that the former made no distinction in both syntax and semantics between intentional states, and between intentional states ascribed by the advisor and held by the advisor. The elementary parts of the semantic domain are no longer the relations for and the propositions, but now, propositions are categorized, what was previously a whole set of propositions is now broken down into several sets. All propositions are now split along two dimensions: a proposition is either assigned to the advisor or ascribed by the advisor to the decision-maker, and a proposition is either then believed, desired, or otherwise.

It is critical to see that every proposition remains with the advisor, as it is his perspective that we are taking. He thus has cues about the decision-maker, and it is on the basis of the intentional state that the advisor ascribes to the decision-maker that a proposition obtains its label. Not all propositions are ascribed to the decision-maker though, some of them are advisor's own, that is, they refer to what the advisor can be said to believe, desire, intend, or how he may evaluate. Hence the distinction then in Fig. 6.8 at the level of the alphabet between signs for propositions which are held by the advisor and the propositions that he ascribes to the decision-maker. Bear in mind, as should be obvious, that all of these propositions are held by the advisor, as when he ascribes something, he must be making assumptions about what the decision-maker may believe, desire, or otherwise.

If there are cues and propositions held by the advisor, where is then advice in AML1? We said in Chap. 5 that advice is about cues, so that advice must be propositions which are *not* already ascribed to the decision-maker, for there would otherwise be no reason to advise at all. It follows that only a labeled proposition which is held by the advisor and not ascribed to the decision-maker can become advice. It is by being communicated that a held labeled proposition becomes advice. Which kind of advice, i.e., advice-about-an-assumption, advice-about-a-goal, and so on, the labeled proposition becomes depends on the kind of cue it targets, as the advice ontology pointed out. The labeled propositions held by the advisor and shown in Fig. 6.8 are consequently *potential* advice, and it is only if those of them which the advisor communicates to the decision-maker which then become advice.



**Fig. 6.8** Alphabet, semantic mapping, and the semantic domain for AML1. There are three sets of signs in the alphabet: (1) signs for connectives, (2) signs for labeled propositions held of the decision-maker, and (3) signs for labeled propositions held by the advisor. The figure shows that some instances of a concept, e.g., *goal* are ascribed by the advisor to the decision-maker and are consequently cues, while other goals are instead held by the advisor. What the figure does not show is that some of the assumptions, goals, evaluations, and commitments held by the advisor may become advice, i.e., be communicated to the decision-maker

To introduce the classification of propositions and expressions in the modeling language itself, we add a function which tells us what concept a given proposition or expression instantiates.

**Definition 6.12. Labeling function for propositions and expressions ( $\mathcal{L}$ ):** Let  $P$  be the set of all propositions and  $E$  the set of all expressions. The function

$$\mathcal{L} : P \cup E \rightarrow \mathcal{L}$$

returns for every proposition and every expression a label from the set

$$\mathcal{L} = \{\mathbf{a}_D, \mathbf{g}_D, \mathbf{e}_D, \mathbf{c}_D, \mathbf{a}_A, \mathbf{g}_A, \mathbf{e}_A, \mathbf{c}_A\}$$

The function is completely defined as follows:

- $\mathcal{L}(p) = \mathbf{a}_D$  if and only if (hereafter iff) the proposition  $p$  is an instance of assumption,
- $\mathcal{L}(\phi) = \mathbf{a}_D$  iff the expression  $\phi$  is an instance of assumption,
- $\mathcal{L}(p) = \mathbf{g}_D$  iff  $p$  is an instance of goal,
- $\mathcal{L}(p) = \mathbf{e}_D$  iff  $p$  is an instance of evaluation,
- $\mathcal{L}(p) = \mathbf{c}_D$  iff  $p$  is an instance of commitment,
- $\mathcal{L}(p) = \mathbf{a}_A$  iff  $p$  is the advisor's assumption (i.e., an assumption that the advisor holds himself),
- $\mathcal{L}(\phi) = \mathbf{a}_A$  iff the expression  $\phi$  is the advisor's assumption,
- $\mathcal{L}(p) = \mathbf{g}_A$  iff  $p$  is the advisor's goal,
- $\mathcal{L}(p) = \mathbf{e}_A$  iff  $p$  is the advisor's evaluation,
- $\mathcal{L}(p) = \mathbf{c}_A$  iff  $p$  is the advisor's commitment.

Table 6.1 gives a synthesis of the rules in Definition 6.12 and can be seen as a top to bottom reading of Fig. 6.8. We also see that it is more precise to say here that we are categorizing propositions and expressions according to the cue ontology, and that some of these cues become advice only if they are communicated by the advisor, rather than that both the cue and advice perform the classification of propositions and expressions.

What effectively happened by moving from AML2 to AML1 is that we took something elementary in the semantic domain of AML2 and tried to see what we can extract from it. We thus took propositions, and looked into what they may be conveying, found that in AML2 they convey both the intentional state (e.g., “*the advisor believes that the Earth is flat*”), whom that intentional state is ascribed (“*the advisor believes that the Earth is flat*”), and the object of the intentional state (“*the advisor believes that the Earth is flat*”). What, then, AML1 does is that it

**Table 6.1** Definition of the labeling function for propositions and expressions ( $\mathcal{L}$ ) in AML1

$y$	INSTANCE OF	$\mathcal{L}(y) =$
$p$	assumption	$\mathbf{a}_D$
$\phi$	assumption	$\mathbf{a}_D$
$p$	goal	$\mathbf{g}_D$
$p$	evaluation	$\mathbf{e}_D$
$p$	commitment	$\mathbf{c}_D$
$p$	A's assumption	$\mathbf{a}_A$
$\phi$	A's assumption	$\mathbf{a}_A$
$p$	A's goal	$\mathbf{g}_A$
$p$	A's evaluation	$\mathbf{e}_A$
$p$	A's commitment	$\mathbf{c}_A$

extracts and highlights the two relations which were otherwise lumped together with the object of the intentional state. One of these relations is that between the individual and the intentional state, hence the subscripts **A** and **D** in the alphabet in AML1. The other is the relationship between the object of the intentional state and the psychological mode, which is why we need the signs **a**, **g**, **e**, **c**. One can then anticipate how AML will compare to the syntax and semantics of AML1: it will look for further relations that in AML1 remain inside propositions, and take them out of these propositions.

The remaining work we should now invest in defining AML1 is not particularly difficult, as it involves the rewriting of the definitions to accommodate that we have labeled propositions and expressions, instead of only propositions and expressions. The inference and conflict relations in AML1 are thus very similar to those in AML2, as the classification we introduce changes nothing in the purpose of these relations within the modeling language.

**Definition 6.13. Inference relation ( $\mathbf{I}$ ).** If the semantic domain includes:

1. the labeled propositions referred to by  $pl_1, \dots, pl_n$ ,
2. the conjunction relation between propositions  $pl_1, \dots, pl_n$ ,
3. the implication relation according to which  $\bigwedge_{i=1}^n pl_i$  implies  $pl$

then the semantic domain also includes the labeled proposition referred to by  $pl$  and it includes the inference relation

$$\frac{pl_1 \quad \dots \quad pl_n \quad \bigwedge_{i=1}^n pl_i \rightarrow pl}{pl}$$

in which the labeled propositions  $pl_1, \dots, pl_n$  and the expression  $\bigwedge_{i=1}^n pl_i \rightarrow pl$  are premises and  $pl$  is the conclusion of the inference relation.

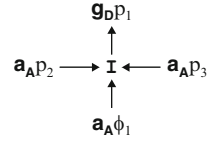
Definition 6.13 is the rewriting of Definition 6.3 to accommodate that propositions and expressions are now labeled.

*Example 6.8.* A political advisor to a prince is keen on Machiavelli's thinking, enthusiastic to apply it to unexpected subjects whenever the occasion presents itself. He is the advisor **A**, and his prince the decision-maker **D** who happens to have inherited principalities. As **A** follows Machiavelli's suggestions, ones from the passage quoted in Example 6.2, he ascribes to **D** a willingness to preserve political power in his inherited principalities. **A** thus has a cue about what **D** desires. This first piece of information is  $\mathbf{g_D} p_1$ , where  $p_1$  abbreviates "to preserve a state which is hereditary and accustomed to the family of their prince," and  $p_1$  is considered as a cue, an instance of goal more specifically which the advisor ascribes to the prince.

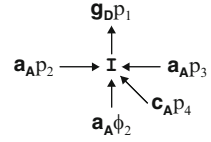
If the advisor assumes that **D** knows not how to preserve his inherited principalities, **A** need not assign further cues to **D**. As **A** follows Machiavelli's reasoning, he will advise what Machiavelli would, that to preserve a state which is hereditary and accustomed to the family of their prince, **D** should "not neglect the constitutional arrangements made by one's predecessors" ( $p_2$ ) and he ought to "adapt one's



**Fig. 6.9** Advisor's reasoning in Example 6.8 shown as a graph, before choosing what advice to give



**Fig. 6.10** Advisor's reasoning in Example 6.8 shown as a graph, after choosing to advise **D** to  $\mathbf{a}_A p_2$ ,  $\mathbf{a}_A p_3$ , and  $\mathbf{a}_A \phi_1$



conduct to circumstances as they arise" ( $p_3$ ), because "if he does that, then he will preserve his inherited principalities" ( $p_2 \wedge p_3 \rightarrow p_1$ ). The advisor assumes all three,  $p_2$ ,  $p_3$ , and  $p_2 \wedge p_3 \rightarrow p_1$ , so that they are labeled as follows:  $\mathbf{a}_A p_2$ ,  $\mathbf{a}_A p_3$ , and  $\mathbf{a}_A \phi_1$ , where  $\phi_1 = \mathbf{a}_A p_2 \wedge \mathbf{a}_A p_3 \rightarrow \mathbf{g}_D p_1$ . Since the advisor concludes  $\mathbf{g}_D p_1$  from  $\mathbf{a}_A p_2$ ,  $\mathbf{a}_A p_3$ , and  $\mathbf{a}_A \phi_1$ , he would advise  $\mathbf{a}_A p_2$ ,  $\mathbf{a}_A p_3$ , and  $\mathbf{a}_A \phi_1$ . Moreover, since  $\mathbf{a}_A p_2$ ,  $\mathbf{a}_A p_3$ , and  $\mathbf{a}_A \phi_1$  lead **A** to conclude  $\mathbf{g}_D p_1$ , they support  $\mathbf{g}_D p_1$  and so each of the  $\mathbf{a}_D p_2$ ,  $\mathbf{a}_D p_3$ , and  $\mathbf{a}_A \phi_1$  becomes, *after being communicated by the advisor*, an instance of advice-about-a-goal. Figure 6.9 illustrates this case, before advice is communicated.

To communicate  $\mathbf{a}_A p_2$ ,  $\mathbf{a}_A p_3$ , and  $\mathbf{a}_A \phi_1$ , **A** ought to commit to advise **D** and thus communicate  $\mathbf{a}_A p_2$ ,  $\mathbf{a}_A p_3$ , and  $\mathbf{a}_A \phi_1$ .  $\mathbf{a}_A p_2$ ,  $\mathbf{a}_A p_3$ , and  $\mathbf{a}_A \phi_1$  consequently become advice only together with the advisor's commitment to communicate them to the decision-maker. If we abbreviate with  $\mathbf{c}_A p_4$  the proposition "Advise  $\mathbf{a}_A p_2$ ,  $\mathbf{a}_A p_3$ , and  $\mathbf{a}_A \phi_1$  to the decision-maker," then the advisor will conclude  $\mathbf{g}_D p_1$  no longer if  $\mathbf{a}_A p_2$ ,  $\mathbf{a}_A p_3$ , and  $\mathbf{a}_A \phi_1$ , but now only if  $\mathbf{a}_D p_2$ ,  $\mathbf{a}_D p_3$ , and  $\mathbf{a}_A \phi_1$ , but if  $\mathbf{a}_A p_2$ ,  $\mathbf{a}_A p_3$ ,  $\mathbf{c}_A p_4$ , and  $\mathbf{a}_A \phi_2$ , where  $\phi_2 \equiv \mathbf{a}_A p_2 \wedge \mathbf{a}_A p_3 \wedge \mathbf{c}_A p_4 \rightarrow \mathbf{g}_D p_1$ . This case is shown in Fig. 6.10.

What we see in this example and in Figs. 6.9–6.10 is that AML1 captures the advice-giving situation in two steps, the first being the reasoning of the advisor before he chooses what to advise, and the second, which shows what the advisor chose to give as advice.

The next step is to replace propositions for labeled propositions in the definition of the conflict relation.

**Definition 6.14. Conflict relation (C).** If the semantic domain includes:

1. The labeled propositions referred to by  $pl_1, \dots, pl_n$
2. The conjunction relation between labeled propositions  $pl_1, \dots, pl_n$
3. The implication relation according to which  $\bigwedge_{i=1}^n pl_i$  implies inconsistency ( $\perp$ ),

then the semantic domain also includes the conflict relation

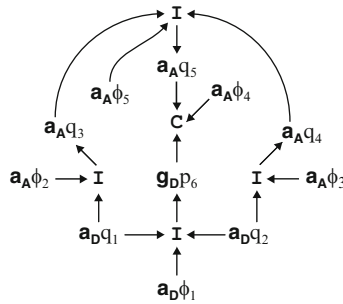
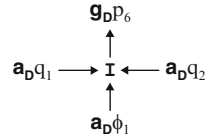
$$\frac{pl_1 \quad \dots \quad pl_n \quad \bigwedge_{i=1}^n pl_i \rightarrow \perp}{\perp}$$

in which the labeled propositions  $pl_1, \dots, pl_n$  and the expression  $\bigwedge_{i=1}^n pl_i \rightarrow \perp$  are premises and  $\perp$  is the conclusion of the conflict relation.

*Example 6.9.* The prince, the decision-maker **D** conquered a principality. The advisor assumes that the prince wishes to “keep the mixed principality without much difficulties and troubles” ( $p_6$ ), i.e., the advisor ascribes a goal to the prince,  $\mathbf{g}_D p_6$ . Moreover, it looks to the advisor as if the prince is convinced that if he is to keep the mixed principality, then he should overthrow a lesser ruler who is round about ( $q_1$ ) and cooperate with a mighty foreign ruler to overthrow other lesser rulers ( $q_2$ ). The advisor ascribes to the prince the rationale shown in Fig. 6.11 which shows that the **A** ascribes  $\mathbf{g}_D p_6$ ,  $\mathbf{a}_D q_1$ ,  $\mathbf{a}_D q_2$ , and  $\mathbf{a}_D q_1 \wedge \mathbf{a}_D q_2 \rightarrow \mathbf{g}_D p_6$  to **D**.

If the advisor read Machiavelli and is enthusiastic about applying what he thus learned, **A** may assume that (1) if the prince overthrows a lesser ruler ( $q_1$ ), then **D** will “lose the trust of other lesser rulers” ( $q_3$ ), i.e.,  $q_1 \rightarrow q_3$ , and that (2) if **D** cooperates with a mighty foreign ruler to overthrow other lesser rulers ( $q_2$ ), then **D** will introduce into the principality that mighty foreign ruler ( $q_4$ ), i.e.,  $q_2 \rightarrow q_4$ . Moreover, the advisor can very well assume – if he further follows Machiavelli – that if the prince loses the trust of other lesser rulers ( $q_3$ ) and introduces a foreign ruler into the principality ( $q_4$ ), then he will lose the principality ( $q_5$ ), i.e.,  $q_3 \wedge q_4 \rightarrow q_5$ . As losing the principality contradicts keeping it, we can write  $p_6 \wedge q_5 \rightarrow \perp$  as well. We can draw this as in Fig. 6.12.

**Fig. 6.11** Rationale that the advisor ascribes to the decision-maker in Example 6.9.  $\phi_1$  abbreviates  $\mathbf{a}_D q_1 \wedge \mathbf{a}_D q_2 \rightarrow \mathbf{g}_D p_6$



**Fig. 6.12** Advisor's rationale in Example 6.9 after he has ascribed the goal  $\mathbf{g}_D p_6$  and the assumptions  $\mathbf{g}_D p_6$ ,  $\mathbf{a}_D q_1$ ,  $\mathbf{a}_D q_2$ , and  $\mathbf{a}_D q_1 \wedge \mathbf{a}_D q_2 \rightarrow \mathbf{g}_D p_6$  to the decision-maker, and if the advisor follows Machiavelli's recommendations. The advisor concludes that the prince will lose the principality if he does as the advisor assumes he will. Abbreviations are as follows:  $\phi_1 \equiv \mathbf{a}_D q_1 \wedge \mathbf{a}_D q_2 \rightarrow \mathbf{g}_D p_6$ ;  $\phi_2 \equiv \mathbf{a}_D q_1 \rightarrow \mathbf{a}_A q_3$ ;  $\phi_3 \equiv \mathbf{a}_D q_2 \rightarrow \mathbf{a}_A q_4$ ;  $\phi_4 \equiv \mathbf{g}_D p_6 \wedge \mathbf{a}_A q_5 \rightarrow \perp$ ;  $\phi_5 \equiv \mathbf{a}_A q_3 \wedge \mathbf{a}_A q_4 \rightarrow \mathbf{a}_A q_5$

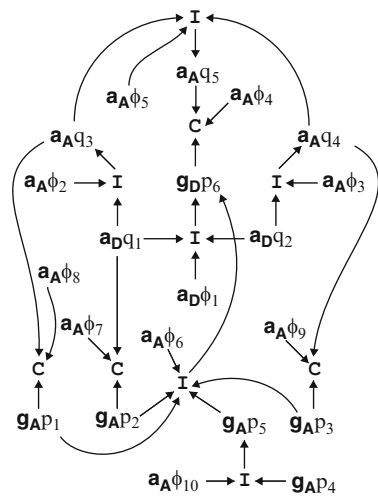
The advisor could choose to give no advice to the decision-maker, thus letting the decision-maker lose the principality. Another option is for the advisor to tell the decision-maker that if **D** goes on according to  $q_1$  and  $q_2$ , then he will lose the principality. In the third option, the advisor recommends courses of action which contradict what **A** assumes that **D** is about to do. Namely, he may suggest that the prince ought to make himself head and defender of lesser rulers who are round about ( $p_1$ ), that he ought to weaken those who are more powerful ( $p_2$ ), take care that no unforeseen event may lead to the appearance of a foreigner as powerful as he is ( $p_3$ ), and finally, the prince should make no effort to win less powerful people, since they are straight away happy to become one with the state he has conquered ( $p_5$ ). If he follows  $p_1$ ,  $p_2$ ,  $p_3$  and  $p_5$ , then he will keep the province, i.e.,  $p_1 \wedge p_2 \wedge p_3 \wedge p_5 \rightarrow p_6$ . Moreover, the prince should know that as soon as he enters a province, all those there who are less powerful come under his influence, moved by the envy which they bear to whoever has been in power over them ( $p_4$ ), and if  $p_4$ , then  $p_5$ , i.e.,  $p_4 \rightarrow p_5$ .

All of this is the same as in Example 6.3, but if the advisor commits to recommend as was just said, then a commitment  $\mathbf{c_A}p_7$  is further needed, which abbreviates “advise  $p_1$ ,  $p_2$ ,  $p_3$ ,  $p_4$ ,  $p_4 \rightarrow p_5$ , and  $p_1 \wedge p_2 \wedge p_3 \wedge p_5 \rightarrow p_6$  to the decision maker,” then we can draw what is being advised as in Fig. 6.13. The upper part of the figure shows what the advisor will be telling to the decision-maker, while the lower part of the figure points to how this advice conflicts with what the advisor ascribes to the decision-maker, namely  $\mathbf{a_D}q_1$  and  $\mathbf{a_D}q_2$ .

The labels on propositions and expressions do not affect the purpose of the consequence relation, picosolution, candidate solution, and maximal candidate solution concepts, so that the three definitions are only slightly modified in AML1 compared to their variants in AML2.

**Fig. 6.13** Advisor chose to recommend to the prince to follow Machiavelli's advice on how to keep his mixed province. Abbreviations are as follows:

- $\phi_2 \equiv \mathbf{a_D}q_1 \rightarrow \mathbf{a_A}q_3$ ;
- $\phi_3 \equiv \mathbf{a_D}q_2 \rightarrow \mathbf{a_A}q_4$ ;
- $\phi_6 \equiv \mathbf{g_A}p_1 \wedge \mathbf{g_A}p_2 \wedge \mathbf{g_A}p_3 \wedge \mathbf{g_A}p_5 \wedge \mathbf{c_A}p_7 \rightarrow \mathbf{g_D}p_6$ ;
- $\phi_7 \equiv \mathbf{a_D}q_1 \wedge \mathbf{g_A}p_2 \rightarrow \perp$ ;
- $\phi_8 \equiv \mathbf{a_D}q_3 \wedge \mathbf{g_A}p_1 \rightarrow \perp$ ;
- $\phi_9 \equiv \mathbf{a_D}q_4 \wedge \mathbf{g_A}p_3 \rightarrow \perp$ ;
- $\phi_{10} \equiv \mathbf{g_D}p_4 \rightarrow \mathbf{g_D}p_5$



**Definition 6.15. Consequence relation ( $\vdash$ ):** Let  $\Gamma$  be some nonempty set of labeled propositions and/or labeled expressions in AML1. Let  $x$  be an abbreviation of either a labeled proposition  $pl$  or of inconsistency, i.e.,  $x \in \{pl, \perp\}$ :

1.  $\Gamma \vdash pl$  if  $pl \in \Gamma$ , or
2.  $\Gamma \vdash x$  if for all  $i$  such that  $1 \leq i \leq n$ ,  $\Gamma \vdash pl_i$  and  $\bigwedge_{i=1}^n pl_i \rightarrow x$  is in  $\Gamma$ , i.e.,  $(\bigwedge_{i=1}^n pl_i \rightarrow x) \in \Gamma$ .

**Definition 6.16. Picosolution for a labeled proposition.** Let  $\mathbf{S}(\mathcal{A})$  be the set of all given labeled propositions and labeled expressions, and  $\mathbf{E}(\mathcal{A})$  the set of all given expressions. A set  $\Gamma$  of labeled propositions and labeled expressions is a picosolution for a labeled proposition  $pl$ , written  $\langle \Gamma, pl \rangle$ , if and only if:

1.  $pl$  is a consequence of  $\Gamma$ :  $\Gamma \vdash pl$ ,
2. every labeled proposition and labeled expression in  $\Gamma$  is a given piece of information:  $\Gamma \subseteq \mathbf{S}(\mathcal{A})$ ,
3. no contradiction is a consequence of  $\Gamma$ :  $\Gamma \not\vdash \perp$ ,
4. all labeled expressions are in  $\Gamma$ :  $\mathbf{E}(\mathcal{A}) \subseteq \Gamma$ ,
5. there are no unnecessary labeled propositions in  $\Gamma$ : there is no  $\Gamma_i \subset \Gamma$  such that both  $\Gamma_i \vdash pl$  and  $\mathbf{E}(\mathcal{A}) \subseteq \Gamma_i$ .

**Definition 6.17. Candidate solution:** A set  $\mathcal{S}$  of picosolutions is a candidate solution if and only if  $\mathcal{S} = \{\langle \Gamma_i, pl_i \rangle \mid \bigcup_{i=1}^n \Gamma_i \not\vdash \perp\}$ .

**Definition 6.18. Maximal candidate solution:** A candidate solution  $\mathcal{S}$  is maximal if and only if there is no other candidate solution  $\mathcal{S}_i$  such that  $\mathcal{S} \subset \mathcal{S}_i$ .

### 6.3.3 AML

Propositions refer to the qualities of, and relations between objects, referring thereby to conditions, events, and situations. Making a modeling language involves deciding which of that information should remain in the propositions and which is interesting enough – with regards to the purpose of the language – to be taken out of propositions and obtain its very own signs in the alphabet, role in grammar, and be distinguished in the semantic domain from the propositions. In AML2 these were the conjunction, implication, inference, and conflict relations. AML1 then took seriously the relations between the proposition and the intentional state, and between the intentional state being held either by the advisor or ascribed to the decision-maker. The overall idea then, in moving from AML2 to AML1 and now to AML is to take increasing chunks of information from propositions and give this information signs, a role in the grammar, and distinguish them from the rest of the semantic domain.

Examples of what we cannot say through relations in AML1 and we will be able to in AML is that what some conditions are more desirable than others, that the decision-maker or advisor is more confident that some events will occur rather than

others, that the advisor has chosen to advise some specific set of goals and assumptions to the decision-maker instead of another set of goals and assumptions. This is done by rewriting some labeled propositions as relations between other propositions. Roughly speaking, when we see that a proposition  $p$  says that another proposition  $q$  stands in the relation  $X$  to some third proposition  $r$ , we will say that  $p$  is an abbreviation of  $qXr$ , while defining  $X$  as a relation in the language in the similar way we did for the inference and conflict relations in both AML2 and AML1. As a rule then, each proposition that conveys a specific relation which we are particularly interested in is rewritten as the relation between these propositions and the relation is defined within the language.

To do the rewriting, we ought evidently have more symbols in the alphabet of AML compared to alphabets of AML1 and AML2. There are ten relations in AML. In addition to conjunction, implication, inference, and conflict encountered already in AML2 and AML1, AML adds six relations with their corresponding signs: confidence ( $\mathbf{a}^u$ ), preference ( $\mathbf{e}^p$ ), mandatory ( $\mathbf{e}^m$ ), optional ( $\mathbf{e}^o$ ), decision ( $\mathbf{c}^d$ ), and advice ( $\mathbf{c}^a$ ). The reason why the signs for the relations resemble those we used to label propositions in AML1 is that the relations were obtained by the specialization of the assumption, evaluation, and commitment concepts. We will discuss these specializations later on, as soon as the alphabet and grammar are out of the way.

**Definition 6.19. AML alphabet:** The alphabet of AML is exactly the following set of signs:

$$\begin{aligned} & \{p, p_1, \dots, p_n, q, q_1, \dots, q_n, \dots\} \cup \{\phi, \phi_1, \dots, \phi_n, \psi, \psi_1, \dots, \psi_n, \dots\} \\ & \cup \{\mathbf{a}_d, \mathbf{g}_d, \mathbf{a}_d^u, \mathbf{e}_d^p, \mathbf{e}_d^m, \mathbf{e}_d^o, \mathbf{c}_d^d, \mathbf{a}_a, \mathbf{g}_a, \mathbf{a}_a^u, \mathbf{e}_a^p, \mathbf{e}_a^m, \mathbf{e}_a^o, \mathbf{c}_a^a\} \\ & \cup \left\{ \wedge, \rightarrow, (, ), \perp, \frac{\text{premises}}{\text{conclusion}} \right\} \end{aligned}$$

where  $n \geq 1$  is some positive natural number.

Comparison of the alphabets of AML and of AML1 shows that Definition 6.19 replaces:

- $\mathbf{a}_d$  with two signs:  $\mathbf{a}_d$  for any assumption which does not convey a confidence relation ascribed to the decision-maker, and  $\mathbf{a}_d^u$  for a confidence relation ascribed to the decision-maker.
- $\mathbf{a}_a$  with two signs:  $\mathbf{a}_a$  for an assumption (which does not convey a confidence relation) held by the advisor, and  $\mathbf{a}_a^u$  for a confidence relation held by the advisor.
- $\mathbf{e}_d$  with three signs:  $\mathbf{e}_d^p$  for a preference relation ascribed to the decision-maker,  $\mathbf{e}_d^m$  for a mandatory relation ascribed to the decision-maker, and  $\mathbf{e}_d^o$  for an optional relation ascribed to the decision-maker.
- $\mathbf{e}_a$  with three signs:  $\mathbf{e}_a^p$  for a preference relation held by the advisor,  $\mathbf{e}_a^m$  for a mandatory relation held by the advisor, and  $\mathbf{e}_a^o$  for an optional relation held by the advisor.
- $\mathbf{c}_d$  with  $\mathbf{c}_d^d$  for a decision/choice ascribed to the decision-maker.
- $\mathbf{c}_a$  with  $\mathbf{c}_a^a$  for a choice of the advisor to give advice to the decision-maker.

The labels have thus changed in AML compared to AML1. While there are still signs for assumptions and goals, some assumptions are singled out (those that convey a confidence relation), while evaluations are replaced with the preference, mandatory, and optional relations, and commitments are either decisions to act in some ways (and following the definition of commitment choices of sets of goals and assumptions) either by the decision-maker or the advisor, or the advisor's decision to give some advice to the decision-maker.

**Definition 6.20. AML grammar:** The labeled propositions and expressions of AML are those obtained by using the rules below, and only those, finitely many times:

- If  $p$  is sign that refers to a proposition, then  $\mathbf{a}_A p$ ,  $\mathbf{g}_A p$ ,  $\mathbf{a}_D p$ ,  $\mathbf{g}_D p$  are labeled propositions.
- If  $pl$  is a sign that refers to a labeled proposition, then  $\mathbf{e}_A^M(pl)$ ,  $\mathbf{e}_A^O(pl)$ ,  $\mathbf{e}_D^M(pl)$ ,  $\mathbf{e}_D^O(pl)$  are labeled unary relations.
- If  $pl_1, pl_2$  are signs that each refers to a labeled proposition, then  $\mathbf{a}_A^U(pl_1, pl_2)$ ,  $\mathbf{e}_A^P(pl_1, pl_2)$ ,  $\mathbf{a}_D^U(pl_1, pl_2)$ ,  $\mathbf{e}_D^P(pl_1, pl_2)$  are labeled binary relations.
- If  $pl_1, \dots, pl_n$  are signs that refer to labeled propositions, labeled unary relations, and/or labeled binary relations, then  $\mathbf{c}_A^A(pl_1, \dots, pl_n)$ ,  $\mathbf{c}_D^D(pl_1, \dots, pl_n)$  are labeled  $n$ -ary relations.
- If  $pl_1, \dots, pl_n$  are signs that refer to labeled propositions, labeled unary relations, labeled binary relations, and/or labeled  $n$ -ary relations, then  $pl_1 \wedge pl_2 \wedge \dots \wedge pl_n \rightarrow pl$  is an expression.
- If  $pl_1, \dots, pl_n$  are signs that refer to labeled propositions, labeled unary relations, labeled binary relations, and/or labeled  $n$ -ary relations, then  $pl_1 \wedge pl_2 \wedge \dots \wedge pl_n \rightarrow \perp$  is an expression.
- If  $\phi$  is an expression, then  $\mathbf{a}_D \phi$  and  $\mathbf{a}_A \phi$  are labeled expressions.

Above,  $n \geq 1$  is some positive natural number.

Definition 6.20 in Backus Naur form is as follows:

$$pl ::= \mathbf{a}_A p \mid \mathbf{g}_A p \mid \mathbf{a}_D p \mid \mathbf{g}_D p, \quad (6.5)$$

$$ur ::= \mathbf{e}_A^M(pl) \mid \mathbf{e}_A^O(pl) \mid \mathbf{e}_D^M(pl) \mid \mathbf{e}_D^O(pl), \quad (6.6)$$

$$br ::= \mathbf{a}_A^U(pl_1, pl_2) \mid \mathbf{e}_A^P(pl_1, pl_2) \mid \mathbf{a}_D^U(pl_1, pl_2) \mid \mathbf{e}_D^P(pl_1, pl_2), \quad (6.7)$$

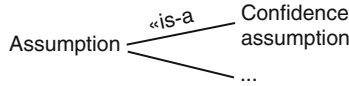
$$\phi ::= \bigwedge_{i=1}^n pr_i \rightarrow pr \mid \bigwedge_{i=1}^n pr_i \rightarrow \perp, \quad (6.8)$$

$$el ::= \mathbf{a}_D \phi \mid \mathbf{a}_A \phi, \quad (6.9)$$

$$pr ::= pl \mid ur \mid br \mid el, \quad (6.10)$$

$$nr ::= \mathbf{c}_A^A(pr_1, \dots, pr_n) \mid \mathbf{c}_D^D(pr_1, \dots, pr_n). \quad (6.11)$$

We have above first assumptions and goals as labeled propositions  $pl$ . Any assumption/goal can have then be mandatory or optional. Any pair of labeled propositions,  $pl_1$  and  $pl_2$ , can be in a confidence or preference relation, as we see



**Fig. 6.14** Confidence assumption concept is a specialization of the Assumption concept. The *three dots* should be read as saying that we are not interested in how to specialize assumptions which are not confidence assumptions; we are only interested in isolating confidence assumptions from all other assumptions

from the line that defines *br*. We then define expressions and label them, before stating that *pr* is either a labeled proposition, a unary relation, a binary relation, or an expression. We can then use *pr* to say in line that defines *nr* that any set of labeled propositions, unary relations, binary relations, and/or expressions can be the object of a commitment by the advisor or the decision-maker.

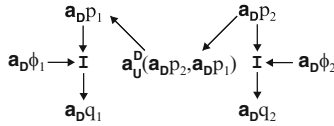
The semantic domain of AML is partitioned onto labeled propositions and relations. We keep the conjunction, implication, inference, and conflict from AML1 as well as the assumptions and goals.

Some assumptions convey an individual's confidence in the occurrence of events, the satisfaction of conditions. Such propositions are a particular kind of assumptions, those which compare in terms of the individual's confidence other assumptions and/or goals. Propositions on confidence are thus just assumptions in AML1, as they are but a specialization of the assumption concept. In AML every assumption that conveys the confidence relation is a relation between other labeled propositions, by analogy to either the conflict or inference relation. The new concept, Confidence assumption, is a specialization of assumption as shown in Fig. 6.14 and is defined as a relation in AML.

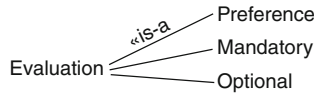
**Definition 6.21. Confidence relation ( $\mathbf{a}^u$ ).** If the semantic domain includes (1) the labeled propositions referred to by  $pl_1$  and  $pl_2$ , (2) an assumption  $\mathbf{a}p$  which states that there is more confidence in  $pl_1$  than in  $pl_2$ , then the semantic domain also includes the binary relation referred to by  $\mathbf{a}^u(pl_1, pl_2)$  and called the confidence relation.

Definition 6.21 simply states that every assumption which states a binary confidence relation between labeled propositions results in the semantic domain in such a relation between the labeled propositions referred to in that assumption. This also has the effect that if  $\mathbf{a}p$  states that one is more confidence in  $pl_1$  than in  $pl_2$ , then  $\mathbf{a}p \equiv \mathbf{a}^u(pl_1, pl_2)$ .

*Example 6.10.* The prince conquered a principality and should decide what to do with the local rulers who surrendered. The advisor could suppose that the prince will take them to be either manageable or not, so  $\mathbf{A}$  ascribes  $\mathbf{a}_D p_1$  and  $\mathbf{a}_D p_2$  to the prince. In  $\mathbf{a}_D p_1$ ,  $p_1$  is for “local rulers are not a threat,” while in  $\mathbf{a}_D p_2$ ,  $p_2$  is for “local rulers are a threat.” The advisor further supposes that if  $p_1$  then the prince will aim to “bring down the rulers” which we denote  $q_1$ , while if  $p_2$  then  $q_2$ , which abbreviates “keep local rulers in place.” The advisor has then also has  $\mathbf{a}_D p_1 \rightarrow \mathbf{g}_D q_1$



**Fig. 6.15** Advisor's rationale in Example 6.10. Abbreviations are:  $\phi_1 \equiv \mathbf{a_D} p_1 \rightarrow \mathbf{a_D} q_1$ ;  $\phi_2 \equiv \mathbf{a_D} p_2 \rightarrow \mathbf{a_D} q_2$



**Fig. 6.16** Evaluation concept specialized onto the preference, mandatory, and optional relations

and  $\mathbf{a_D} p_2 \rightarrow \mathbf{g_D} q_2$ . Knowing the prince, the advisor may further suppose that he is particularly careful to threats; so it is more likely for him that the local rulers are a threat than its opposite. The advisor thus ascribes a confidence assumption to the prince:  $\mathbf{a_U^D}(\mathbf{a_D} p_2, \mathbf{a_D} p_1)$ . Figure 6.15 shows the advisor's rationale.

Evaluations state either a preference, a mandatory relation, or an optional relation on goals and assumptions. Figure 6.16 gives the specialization of the evaluation concept.

Preferences arise out of evaluations that compare two labeled propositions in terms of desirability. As each labeled proposition refers to conditions, events, and situations, it follows that a preference states which of the conditions, events, and situations are preferred by saying which of the two propositions are preferred.

**Definition 6.22. Preference relation ( $\mathbf{e^P}$ ).** If the semantic domain includes (1) the labeled propositions referred to by  $pl_1$  and  $pl_2$ , (2) an evaluation  $\mathbf{e^P}$  which states that  $pl_1$  is strictly more desirable than  $pl_2$ , then the semantic domain also includes the binary relation referred to by  $\mathbf{e^P}(pl_1, pl_2)$  and called the preference relation.

*Example 6.11.* We keep the advisor's rationale from Example 6.10. If the advisor prefers that the decision-maker assumes that the rulers of the conquered lands are not a threat over them being a threat, then the advisor has a preference for  $\mathbf{a_D} p_1$  over  $\mathbf{a_D} p_2$ , i.e.,  $\mathbf{e_A^P}(\mathbf{a_D} p_1, \mathbf{a_D} p_2)$ , as in Fig. 6.17.

Individual propositions can be evaluated in terms of desirability, but independently of other propositions. As we said earlier, an instance of evaluation can thus indicate that a labeled proposition is mandatory or optional. The mandatory relation is a unary relation, instance of evaluation which states that a labeled proposition must be part of a candidate solution to the advisor's problem. That a labeled proposition ascribed to the decision-maker is mandatory thus means in practice that the advisor cannot give advice which contradicts that labeled proposition.

**Definition 6.23. Mandatory relation ( $\mathbf{e^M}$ ).** If the semantic domain includes (1) the labeled proposition referred to by  $pl$ , (2) an evaluation  $\mathbf{e^P}$  which states that  $pl$  must

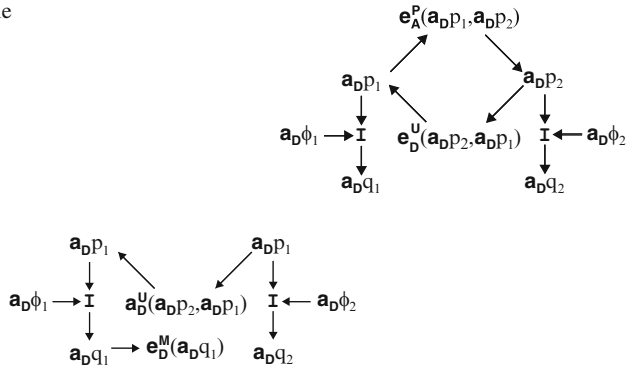


**Fig. 6.17** Advisor's rationale in Example 6.11.

Abbreviations are:

$\phi_1 \equiv \mathbf{a_D} p_1 \rightarrow \mathbf{a_D} q_1$ ;

$\phi_2 \equiv \mathbf{a_D} p_2 \rightarrow \mathbf{a_D} q_2$



**Fig. 6.18** Advisor's rationale in Example 6.12. Abbreviations are:  $\phi_1 \equiv \mathbf{a_D} p_1 \rightarrow \mathbf{a_D} q_1$ ;  $\phi_2 \equiv \mathbf{a_D} p_2 \rightarrow \mathbf{a_D} q_2$

not be countered by advice, then the semantic domain also includes the unary relation referred to by  $\mathbf{e^M}(pl)$  and called the mandatory relation.

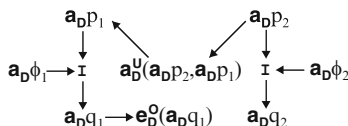
*Example 6.12.* We take again the advisor's rationale from Example 6.10. If the advisor believes that the prince considers a must to take down the rulers of the newly conquered lands, then the advisor ascribes a mandatory relation  $\mathbf{e_D^M}(a_{Dq1})$  to the prince, as shown in Fig. 6.18. The effect of taking  $\mathbf{a_{Dq1}}$  as mandatory is that the advisor will not choose to give to the prince the advice which together with  $\mathbf{a_{Dq1}}$  results in contradiction.

**Definition 6.24. Optional relation ( $\mathbf{e^O}$ ).** If the semantic domain includes (1) the labeled proposition referred to by  $pl$ , (2) an evaluation  $\mathbf{e}p$  which states that it would be preferred if  $pl$  was not countered by advice then if it was, then the semantic domain also includes the unary relation referred to by  $\mathbf{e^O}(pl)$  and called the optional relation.

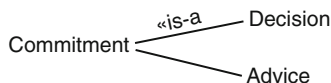
*Example 6.13.* For a careless prince, it may not be critical to remove threatening local rulers, which from the advisor's perspective amounts to  $\mathbf{a_{Dq1}}$  being optional, as in Fig. 6.19. The effect of making  $\mathbf{a_{Dq1}}$  optional is that the advisor could give to the prince the advice which together with  $\mathbf{a_{Dq1}}$  results in contradiction, but that the prince would find it more desirable if the advice did not contradict  $\mathbf{a_{Dq1}}$ .

Commitments of interest tell us either that the advisor ascribed a decision to the decision-maker, or what the advisor chose to give as advice to the decision-maker. The commitment concept gives us decision and ADVICE relations, shown in Fig. 6.20.

Ascribing a commitment to the decision-maker amounts to ascribe him the picking out of some assumptions and goals among others. Of course, both those that are picked out and others are all ascribed by the advisor, but since we said above that contradictions/conflicts are to be avoided, the decision-maker will assume that once there are conflicts, and thereby incompatible sets of cues, the decision-maker will



**Fig. 6.19** Advisor's rationale in Example 6.13. Abbreviations are:  $\phi_1 \equiv \mathbf{a}_D p_1 \rightarrow \mathbf{a}_D q_1$ ;  $\phi_2 \equiv \mathbf{a}_D p_2 \rightarrow \mathbf{a}_D q_2$



**Fig. 6.20** Commitment concept specialized onto the decision and advise relations

chose one of them – a decision relation states which among the ascribed cues is chosen by the decision-maker. We argued earlier, in discussing the commitment concept, that ascribing plans to the decision-maker requires that assumptions-about-actions, goals, and commitments to perform some of the actions be ascribed: a decision selects a set of cues, it can select those that may form a plan.

**Definition 6.25. Decide ( $\mathbf{c}_D^D$ ).** If the semantic domain includes (1) cues, unary relations, and binary relations, referred to as  $pr_1, \dots, pr_n$ , (2) a commitment cue  $\mathbf{c}_D p$  which states that the decision-maker chooses  $pr_1, \dots, pr_n$ , then the semantic domain also includes the n-ary relation referred to by  $\mathbf{c}_D^D(pr_1, \dots, pr_n)$  and called the decide relation.

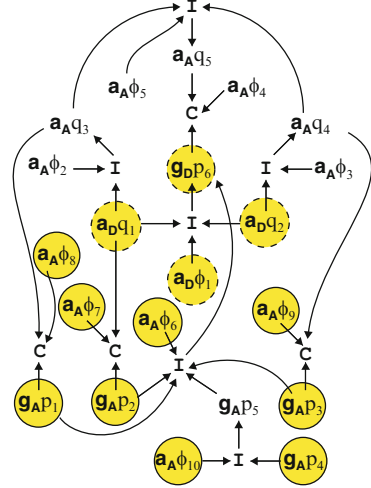
The purpose of the advise relation is to select not the cues of the decision-maker, but the information that the advisor choses to communicate to the decision-maker, the advice he commits to give.

**Definition 6.26. Advise ( $\mathbf{c}_A^A$ ).** If the semantic domain includes (1) cues, unary relations, and binary relations, referred to by  $pr_1, \dots, pr_n$ , (2) a commitment cue  $\mathbf{c}_D p$  which states that the advisor chooses to communicate  $pr_1, \dots, pr_n$  to the decision-maker, then the semantic domain also includes the n-ary relation referred to by  $\mathbf{c}_A^A(pr_1, \dots, pr_n)$  and called the advise relation.

Definition 6.26 gives a somewhat different conception of advice than our discussion in Chap. 3. Departures are minimal, and it is mainly the presentation that changed. Emphasis was on communication and reference in Chap. 3, now it is on the categories of information in advice, namely, the assumptions, goals, and so on, which the advisor chooses to communicate.

*Example 6.14.* In Example 6.9, we said that there was a prince who conquered a principality, and that his advisor assumed that the prince believes that he will keep the principality without much difficulties and troubles ( $\mathbf{g}_D p_6$ ) if he overthrows a lesser ruler who is round about ( $\mathbf{a}_D q_1$ ) and cooperate with a mighty foreign ruler to overthrow other lesser rulers ( $\mathbf{a}_D q_2$ ). The advisor consequently ascribes to the

**Fig. 6.21** The decision that the advisor ascribes to the decision-maker is shown by encircling nodes with a dashed line. Nodes given as advice are encircled by a solid line. For abbreviations, see Fig. 6.13



prince the rationale that we showed in Fig. 6.11. We then considered why the advisor might disagree, if he is a fan of Machiavelli. This led us finally to Fig. 6.13, which shows the conflicts between what the advisor may recommend and the cues he ascribed to the prince. If the advisor ascribes to the prince the decision to keep the principality by overthrowing lesser rulers and cooperating with a foreign ruler, we can show this by adding the decision  $c_D^D(g_D p_6, a_D \phi_1, a_D \phi_2)$ ; the corresponding nodes are encircled with a dashed line in Fig. 6.21.

The advisor may choose to recommend what Machiavelli would suggest, namely,  $c_A^A(g_A p_1, g_A p_2, g_A p_3, g_A p_4, a_A \phi_{10}, a_A \phi_6)$ . The nodes selected by this advice are encircled with a solid line in Fig. 6.21.

The inference and conflict relations in AML remain identical to those of AML1.

**Definition 6.27. Inference relation (I).** If the semantic domain includes:

1. the labeled propositions referred to by  $pl_1, \dots, pl_n$ ,
2. the conjunction relation between propositions  $pl_1, \dots, pl_n$ ,
3. the implication relation according to which  $\bigwedge_{i=1}^n pl_i$  implies  $pl$

then the semantic domain also includes the labeled proposition referred to by  $pl$  and it includes the inference relation

$$\frac{pl_1 \quad \dots \quad pl_n \quad \bigwedge_{i=1}^n pl_i \rightarrow pl}{pl}$$

in which the labeled propositions  $pl_1, \dots, pl_n$  and the expression  $\bigwedge_{i=1}^n pl_i \rightarrow pl$  are premises and  $pl$  is the conclusion of the inference relation.

**Definition 6.28. Conflict relation (C).** If the semantic domain includes:

1. the labeled propositions referred to by  $pl_1, \dots, pl_n$ ,

2. the conjunction relation between labeled propositions  $pl_1, \dots, pl_n$ ,
3. the implication relation according to which  $\bigwedge_{i=1}^n pl_i$  implies inconsistency ( $\perp$ )

then the semantic domain also includes the conflict relation

$$\frac{pl_1 \quad \dots \quad pl_n \quad \bigwedge_{i=1}^n pl_i \rightarrow \perp}{\perp}$$

in which the labeled propositions  $pl_1, \dots, pl_n$  and the expression  $\bigwedge_{i=1}^n pl_i \rightarrow \perp$  are premises and  $\perp$  is the conclusion of the conflict relation.

The inference and conflict relations are defined using *modus ponens*, the only available inference rule in AML. The confidence, preference, mandatory, optional, decision, and advise relations are not defined in terms of inference steps; the consequence is that they do not influence the definition of the consequence relation, picosolution, candidate solution, and maximal candidate solution concepts. The definitions of these notions are identical in AML1 and in AML.

**Definition 6.29. Consequence relation ( $\vdash$ ):** Let  $\Gamma$  be some nonempty set of labeled propositions and/or labeled expressions in AML. Let  $x$  be an abbreviation of either a labeled proposition  $pl$  or of inconsistency, i.e.,  $x \in \{pl, \perp\}$ :

1.  $\Gamma \vdash pl$  if  $pl \in \Gamma$ , or
2.  $\Gamma \vdash x$  if for all  $i$  such that  $1 \leq i \leq n$ ,  $\Gamma \vdash pl_i$  and  $\bigwedge_{i=1}^n pl_i \rightarrow x$  is in  $\Gamma$ , i.e.,  $(\bigwedge_{i=1}^n pl_i \rightarrow x) \in \Gamma$ .

**Definition 6.30. Picosolution for a labeled proposition.** Let  $\mathbf{S}(\mathcal{A})$  be the set of all given labeled propositions and labeled expressions, and  $\mathbf{E}(\mathcal{A})$  the set of all given expressions. A set  $\Gamma$  of labeled propositions and labeled expressions is a picosolution for a labeled proposition  $pl$ , written  $\langle \Gamma, pl \rangle$ , if and only if:

1.  $pl$  is a consequence of  $\Gamma$ :  $\Gamma \vdash pl$ ,
2. every labeled proposition and labeled expression in  $\Gamma$  is a given piece of information:  $\Gamma \subseteq \mathbf{S}(\mathcal{A})$ ,
3. no contradiction is a consequence of  $\Gamma$ :  $\Gamma \not\vdash \perp$ ,
4. all labeled expressions are in  $\Gamma$ :  $\mathbf{E}(\mathcal{A}) \subseteq \Gamma$ ,
5. there are no unnecessary labeled propositions in  $\Gamma$ : there is no  $\Gamma_i \subset \Gamma$  such that both  $\Gamma_i \vdash pl$  and  $\mathbf{E}(\mathcal{A}) \subseteq \Gamma_i$ .

The definitions of the candidate solution and its maximal variant cannot remain the same as in AML1: in addition to being consistent, a set of picosolutions must now include all mandatory nodes. This simply reflects the definition of the mandatory relation, in that a mandatory node must not be contradicted.

**Definition 6.31. Candidate solution:** A set  $\mathcal{S}$  of picosolutions is a candidate solution if and only if:

1.  $\mathcal{S} = \{\langle \Gamma_i, pl_i \rangle \mid \bigcup_{i=1}^n \Gamma_i \not\vdash \perp\}$ , and
2.  $\mathcal{S}$  contains picosolutions for every mandatory labeled proposition.

**Definition 6.32. Maximal candidate solution:** A candidate solution  $\mathcal{S}$  is maximal if and only if there is no other candidate solution  $\mathcal{S}_i$  such that  $\mathcal{S} \subset \mathcal{S}_i$ .

### 6.3.4 A-nets

The graphs that illustrated the examples up to this point are called a-nets. They are simply expressions written in a different syntax, whereby that syntax can be rather straightforwardly defined for each of the three modeling languages. An a-net is written in what we call the *graph syntax*, whereas the alphabets and grammars introduced earlier formed *symbolic syntaxes* for each language.

#### A-nets in AML2

**Definition 6.33. Alphabet for graph syntax in AML2:** The alphabet for the graph syntax of AML2 is exactly the following set of signs:

$$\begin{aligned} &\{p, p_1, \dots, p_n, q, q_1, \dots, q_n, \dots\} \\ &\quad \cup \{\phi, \phi_1, \dots, \phi_n, \psi, \psi_1, \dots, \psi_n, \dots\} \\ &\quad \cup \{\wedge, \rightarrow, (, ), \mathbf{I}, \mathbf{C}, \longrightarrow\}, \end{aligned}$$

where  $n \geq 1$  is some positive natural number.

**Definition 6.34. Grammar for graph syntax in AML2:** An a-net in AML2 is exactly a finite set containing any one or more of the elements shown in Column (a) of Figs. 6.22 and 6.23.

#### A-nets in AML1

**Definition 6.35. Alphabet for graph syntax in AML1:** The alphabet for the graph syntax of AML1 is exactly the following set of signs:

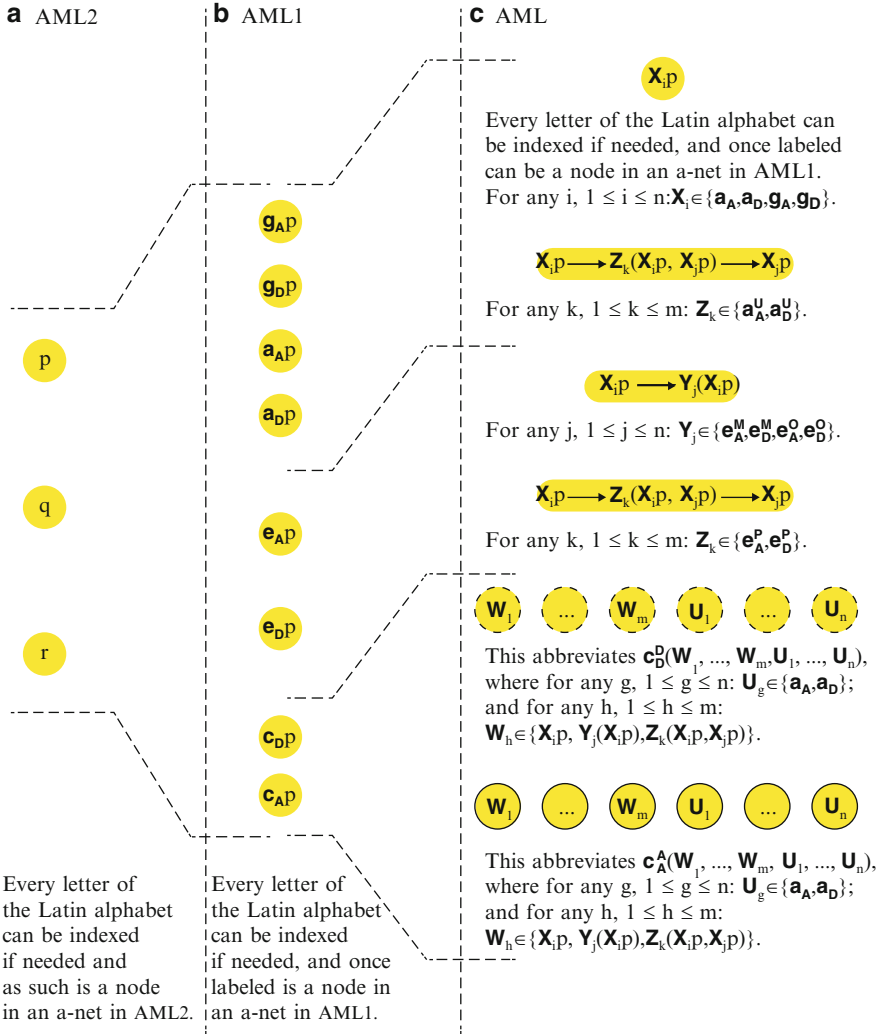
$$\begin{aligned} &\{p, p_1, \dots, p_n, q, q_1, \dots, q_n, \dots\} \cup \{\phi, \phi_1, \dots, \phi_n, \psi, \psi_1, \dots, \psi_n, \dots\} \\ &\quad \cup \{\mathbf{a}_D, \mathbf{g}_D, \mathbf{e}_D, \mathbf{c}_D, \mathbf{a}_A, \mathbf{g}_A, \mathbf{e}_A, \mathbf{c}_A\} \cup \{\wedge, \rightarrow, (, ), \mathbf{I}, \mathbf{C}, \longrightarrow\}, \end{aligned}$$

where  $n \geq 1$  is some positive natural number.

**Definition 6.36. Grammar for graph syntax in AML1:** An a-net in AML1 is exactly a finite set containing any one or more of the elements shown in Column (b) of Figs. 6.22 and 6.23.

#### A-nets in AML

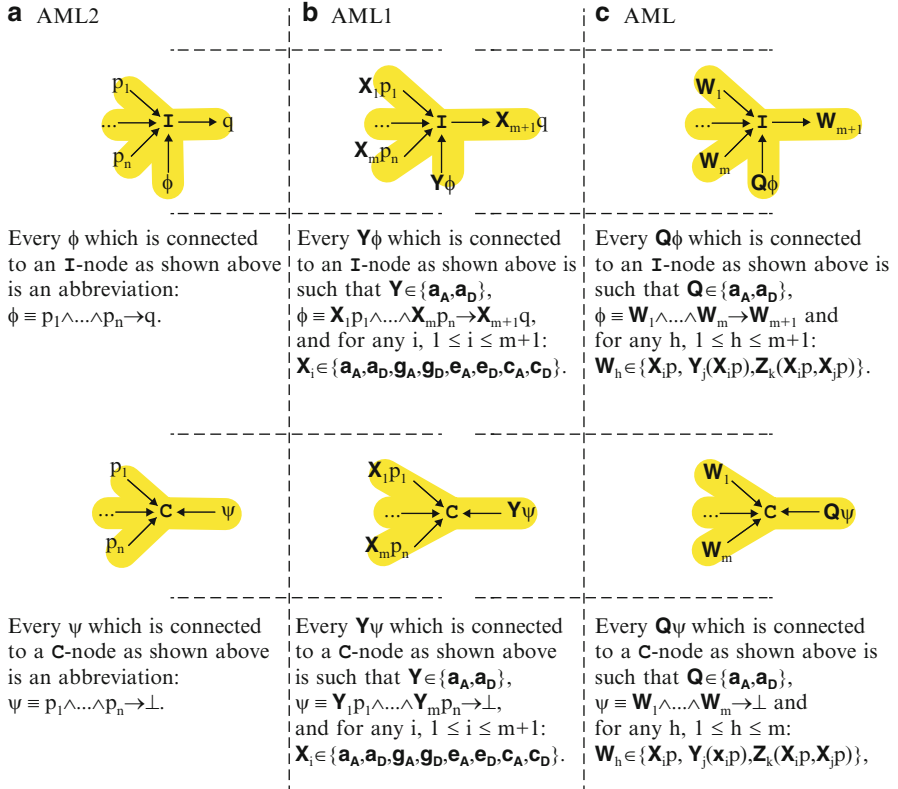
**Definition 6.37. Alphabet for graph syntax in AML:** The alphabet for the graph syntax of AML is exactly the following set of signs:



**Fig. 6.22** Grammars for a-nets in AML2, AML1, and AML. Every a-net in a language of choice is exactly a finite set, each member of which is one of the graphical elements shown in the relevant column: an a-net in AML2 is made of elements in Column (a) in this figure and Fig. 6.23; in AML1 from those in Column (b) in this figure and Fig. 6.23; in AML from those in Column (c) in this figure and Fig. 6.23. Every  $n$  and  $m$  in this figure and Fig. 6.23 are positive natural numbers. The role of dashed lines is explained in the text

$$\begin{aligned}
 & \{p, p_1, \dots, p_n, q, q_1, \dots, q_n, \dots\} \cup \{\phi, \phi_1, \dots, \phi_n, \psi, \psi_1, \dots, \psi_n, \dots\} \\
 & \cup \{a_D, g_D, a_D^u, e_D^p, e_D^m, e_D^o, c_D^p, a_A, g_A, a_A^u, e_A^p, e_A^m, e_A^o, c_A^p\} \\
 & \cup \{\wedge, \rightarrow, (, ), \mathbf{I}, \mathbf{C}, \longrightarrow\},
 \end{aligned}$$

where  $n \geq 1$  is some positive natural number.



**Fig. 6.23** (Continued from Fig. 6.22.) Parts of grammars for a-nets in AML2, AML1, and AML. Every a-net in a language of choice is exactly a finite set, each member of which is one of the graphical elements shown in the relevant column: an a-net in AML2 is made of elements in Column (a) in this figure and Fig. 6.22; in AML1 from those in Column (b) in this figure and Fig. 6.22; in AML from those in Column (c) in this figure and Fig. 6.22. Every  $n$  and  $m$  in the figure are positive natural numbers. The role of dashed lines is explained in the text

**Definition 6.38. Grammar for graph syntax in AML1:** An a-net in AML1 is exactly a finite set containing any one or more of the elements shown in Column (c) of Figs. 6.22 and 6.23.

Figures 6.22 and 6.23 give the grammars for a-nets in the three languages. The dashed lines show how the information referred to by some signs in one language gets referred to via the signs of another language. For example, a proposition  $p$  in AML2 gets labeled in AML1, and we know from the preceding discussions that the label it obtains depends on which cue it instantiates. A labeled proposition in AML1 may change its shape in AML, e.g., when it refers to a confidence, preference, mandatory, optional, decision, or advise relation. The dashed lines also illustrate how we looked into the detail of propositions as we moved from one language to another: it was enough to have propositions in AML2, and we then labeled them





in politics, but for ways in which he presents and defends his case. Choosing which assumptions and goals to advance and which to remain silent on is perhaps equally important as deciding how to form the intended reference relations, which signs to choose in communication and in which context to deliver them. What becomes visible is prominently so, while the effort to reveal what shadows hide deters many. In *Tristes Tropiques*, his autobiography, Claude Lévi-Strauss complains about how he was trained to argue at the Sorbonne of his time:

First you establish the traditional 'two views' of the question. You then put forward a common-sensical justification of the one, only to refute it by the other. Finally, you send them both packing by the use of a third interpretation, in which both the others are shown to be equally unsatisfactory. Certain verbal maneuvers enable you to line up the traditional 'antitheses' as complementary aspects of a single reality: form and substance, content and container, appearance and reality, essence and existence, continuity and discontinuity, and so on. Before long the exercise becomes the merest verbalizing, reflection gives place to a kind of superior punning, and the 'accomplished philosopher' may be recognized by the ingenuity with which he makes ever-bolder play with assonance, ambiguity, and the use of those words which sound alike and yet bear quite different meanings. (Lévi-Strauss 1961, p. 54)

It is, in a manner of speaking, a dance around the claims to the point of impressing by the dance itself, rather than the claims. It consumes attention, and by doing so makes the suggested assumptions and goals seem indubitable, or at least not entirely unacceptable. By the inclusion of countering claims, especially those that are easy to discredit, easy enemies are targeted and can be quickly taken down. The truly difficult ones may best be left unnamed. Once learned, the strategy is easy to apply to whatever issue, and when perfected becomes indistinguishable from bluff.

The interest in structure – in the inference, confidence, and evaluation relations – is a response to such methods. It is an interest in what goes into explanation and prediction, and in how such inputs are related. By illustrating, however simplistically the relations, an a-net hopefully strips the assumptions and goals of theatrics and of verbalizing.

What is striking in an a-net are not so much the lines and nodes, as the white space. It should rather be called the dark space, in spite of the paper's color, for filling it up asks us to uncover what the explanation or prediction remained silent on. Its emptiness refers to what may have escaped the scarce attention, the rationale that may otherwise have doubted the assumptions, goals, conclusions drawn, and comparisons made. To fill it up is difficult, especially if the explanation is carefully crafted. When that is the case, it is the knowledge of alternative explanations that is needed. A literal reading of Machiavelli would thus give clear primacy to resort to military, except when one is familiar with, say, Gene Sharp's politics of nonviolent action.

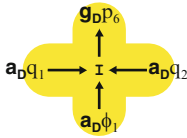
When pictures of explanations and predictions – the a-nets – reveal just how simple, if not simplistic the rationale for choice and advice is, they already fully justify their existence by demystifying what may otherwise appear to be convincing, clear, and efficient solutions.

## 6.4 Formulation of the Advisor's Problem

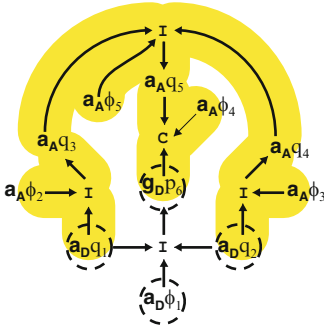
A rough formulation of the problem appeared at the outset of this chapter: given an advisor's explanations and predictions of a decision-maker, what is the advice that this former should give to the latter to influence the predicted choice of the decision-maker? It was not clear at that time what explanations and predictions may look like and what they refer to. Having decision and advise relations, AML defines these notions and relates them to other information relevant the advisor's problem and its resolution.

For a better grasp of how AML helps in formulating the advisor's problem, we return once again to Example 6.9. Figure 6.25 summarizes the information ascribed to the decision-maker, the prince, and held by his advisor. That figure should be read from top to bottom, and understood as showing three steps which we said the advisor may take to determine what to recommend to the prince. The right-hand part of the figure gives the abbreviations used in the a-nets. The left-hand part of the same figure shows the cues which reflect the intentional states that the advisor may have ascribed to the prince as the result of the former's observation of the latter's actions, of their discussions, or whatever else way the advisor may have taken to hypothesize what the prince may be believing, desiring, and so on. We see that the advisor supposes that the prince does want to keep the province, and seems to already have some ideas about how to do so. In this first step, shown in Fig. 6.25a, the advisor in a way considers the cues in isolation, and as our a-net shows, independently from the assumptions, goals, and otherwise, which the advisor may himself hold.

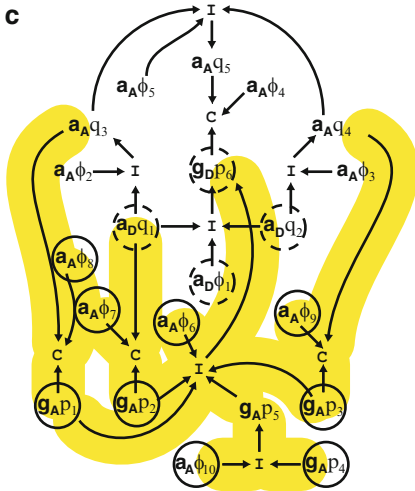
The second step, in Fig. 6.25b, shows the advisor's rationale after taking out these cues from isolation and confronting them to the advisor's own assumptions. Figure 6.25b shows that there are contradictions: the advisor's assumptions disagree with the cues he holds about the decision-maker. If the advisor ascribes to the prince the commitment to act according to these cues, then the advisor supposes that the decision-maker decided to act as these cues suggest. *It is in Fig. 6.25b that we see the problem that the advisor faces in this specific setting: namely, what should the advisor recommend to the decision-maker to detract him from acting as the advisor supposes the decision-maker will act?* It is of course not always the case that advice should go against the supposed decisions of the decision-maker. It need not always aim to detract a hypothesized choice: the advisor may well agree with the choice he ascribes to the decision-maker, and either fully abstain from giving advice, or choose to give recommendations which would support and thereby hopefully reinforce the expected choices of the decision-maker. One response to the advisor's problem is shown in Fig. 6.25c, where the advisor draws on Machiavelli to formulate recommendations. If the advisor chooses to communicate those assumptions and goals, then his choice is drawn by encircling with solid lines the relevant nodes of the a-net, as in Fig. 6.25c. That figure shows one maximal candidate solution to the advisor's problem we just stated. Going back to the rough terms we used to state the advisor's problem at the start of this chapter, we see in this example that the advisor forms explanations and predictions of the decision-maker's behavior by relating cues (as in Fig. 6.25a and Fig. 6.25b), then seeks the information that he

**a***Abbreviations:*

- $p_6$ : Keep the mixed principality without much difficulties and troubles.
- $q_1$ : Overthrow a lesser ruler who is round and about.
- $q_2$ : Cooperate with a mighty foreign ruler to overthrow other lesser rulers.

**b**

- $q_3$ : Lose the trust of other lesser rulers.
- $q_4$ : Introduce a mighty foreign ruler in the principality.
- $q_5$ : Lose the principality.

**c**

- $p_1$ : Make oneself head and defender of lesser rulers who are round about.
- $p_2$ : Weaken those who are more powerful.
- $p_3$ : Take care that no unforeseen event may lead to the appearance of a foreigner as powerful as the prince is.
- $p_4$ : As soon as the prince enters a province, all those there who are less powerful come under his influence, moved by the envy which they bear to whoever has been in power over them.
- $p_5$ : Make no effort to win less powerful people, since they are straight away happy to become one with the state he has conquered.

**Fig. 6.25** (a) The cues the advisor holds about the decision-maker, the prince. The advisor confronts these cues to the assumptions he holds himself, as drawn in (b). (c) Further assumptions that the advisor confronts to the cues, and which could be communicated to the prince so as to change his mind about what to do to keep the province. Differences between each a-net and the one that precedes it in the sequence of figures are highlighted

might choose to communicate to the decision-maker to influence his choice (as in Fig. 6.25c).

If we take into account all the picosolutions from the nodes selected as advice, we obtain a candidate solution. Stated otherwise, if we take the closure of the advice, the result is a candidate solution. It turns out that this candidate solution is also a maximal one in the example shown in Fig. 6.25.

The example made no mention of the confidence, preference, mandatory, and optional relations. They serve for the comparison of candidate solutions – they are the criteria for comparison – as long as there two or more of the candidates. To illustrate their use for comparison, we continue with the example above.

Machiavelli's motives for writing *The Prince* remain unknown, and he is absent to clarify. Perhaps he was sincere, making the case for ruthlessness and deceit, or was he a model citizen intent on leaking the rulers' strategies to the people, supporting thereby dissent. Whatever and regardless of his true motives, his advice is no immutable law of nature making it reasonable to assume that our advisor above may well consider options other than those in *The Prince*. He may deem it less likely that the prince would make himself head and defender of lesser rulers, which results in the confidence relation  $\mathbf{a}_D^U(\mathbf{a}_D q_1, \mathbf{g}_A p_1)$  from the less unlikely  $\mathbf{a}_D q_1$  to the more unlikely  $\mathbf{g}_A p_1$ , as in Fig. 6.26a. The advisor may then choose the easy way out, and simply recommend that the prince does as the advisor already assumed that the prince will do: i.e., the advisor chooses to recommend  $\mathbf{a}_D q_1$ ,  $\mathbf{a}_D q_2$ , and  $\mathbf{a}_D \phi_1$ , as in Fig. 6.26b.

If the advisor is willing to avoid stating the obvious, recommending what he deems the prince already knows, he may in a war-mongering mood offer arguments to reinforce the prince's presumed choice. The advice in this case supports the prince's choice, targeting the cues included in that decision. If the advisor further happened to come across Shakespeare's *Henry V*, he could draw on the advice given to Henry V to reassure him that going to war in France is the right choice (Shakespeare 1914, Act I. Scene II):

*Bishop of Ely:*

Awake remembrance of these valiant dead,  
And with your puissant arm renew their feats:  
You are their heir, you sit upon their throne,  
The blood and courage that renowned them  
Runs in your veins; and my thrice-puissant liege  
Is in the very May-morn of his youth,  
Ripe for exploits and mighty enterprises.

*Duke of Exeter:*

Your brother kings and monarchs of the earth  
Do all expect that you should rouse yourself,  
As did the former lions of your blood.

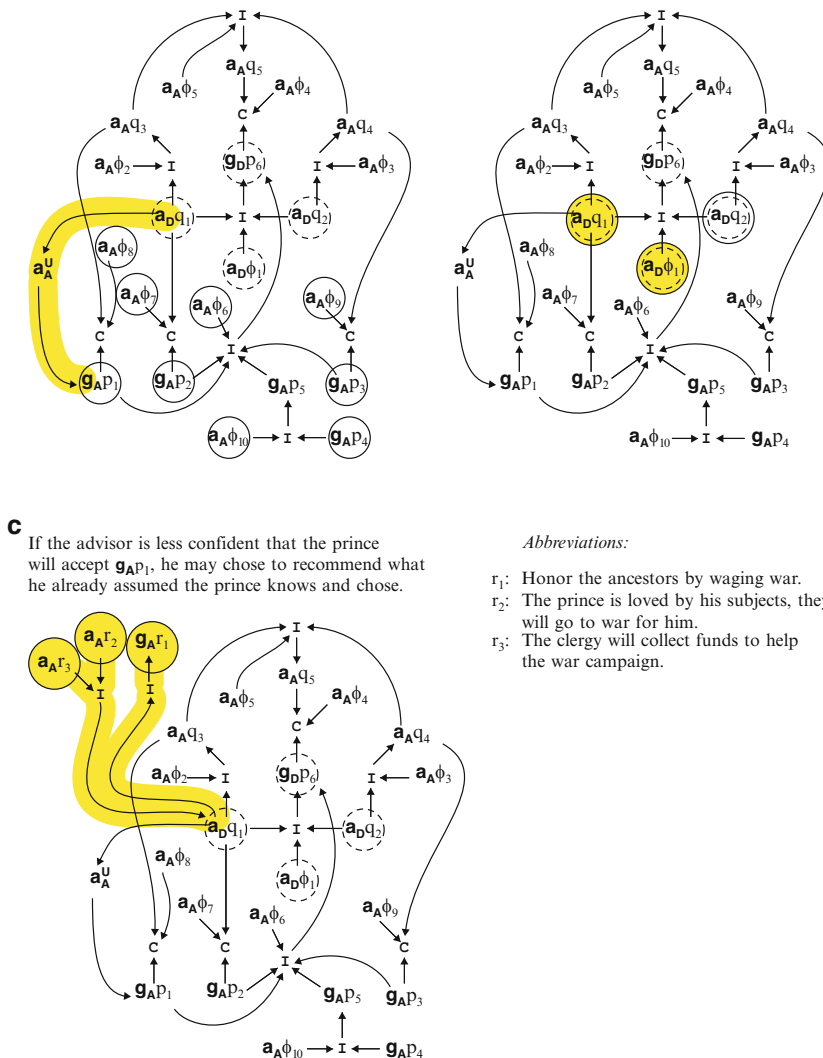
*Earl of Westmoreland:*

They know your Grace hath cause and means and might;  
So hath your highness; never King of England  
Had nobles richer, and more loyal subjects,  
Whose hearts have left their bodies here in England  
And lie pavilion'd in the fields of France.

*Archbishop of Canterbury:*

O! let their bodies follow, my dear liege,  
With blood and sword and fire to win your right;  
In aid whereof we of the spirituality  
Will raise your highness such a mighty sum  
As never did the clergy at one time  
Bring in to any of your ancestors.

- a** A confidence relation, to point out that the advisor is less confident that the prince will accept  $g_{AP1}$  than he is that the prince will follow  $a_{DQ1}$ .
- b** If the advisor is less confident that the prince will accept  $g_{AP1}$ , he may chose to recommend what he already assumed the prince knows and chose.



**Fig. 6.26** Figure 6.26a shows the a-net used in earlier examples, which now also includes a confidence relation. The confidence relation may lead the advisor to reconsider which recommendations he makes: Figure 6.26b and Fig. 6.26c show two candidates. Highlights show differences between a-nets

If we focus on the arguments of the characters, we see that the Bishop of Ely argues the king should wage war to honor his ancestors, while the Duke of Exeter echoes this aim. Earl of Westmoreland is less romantic, suggesting that the king is in a rather favorable position to actually go to war as no other King of England was more loved than he is. The Bishop of Canterbury stays on the practical matters to emphasize less how helpful the love of the subjects may be, but more the clergy's ability to raise funds to help the cause. These same arguments can be given to support the prince interested in overthrowing lesser rulers in his principality, as in Fig. 6.26c.

The confidence relation in Fig. 6.26 taken alone is a criterion for the comparison of alternative advice, and thus of candidate solutions: if we compare the advice in Fig. 6.25c and that in Fig. 6.26c, it is obvious that the former advice includes information which is less likely to be accepted than that included in the latter advice. The comparison here is over a single criterion, as one confidence relation is given, and no preference, mandatory, or optional relations. If the advisor believes that it is critical for the prince to keep the province, i.e., to satisfy the goal  $g_D p_6$ , then the mandatory relation can be added. We can make picosolutions for  $g_D p_6$  from the advice in the a-net in Fig. 6.25c and each a-net in Fig. 6.26b, c. It follows that any of the three alternative ways to advise gives us a candidate solution.

However, if, say,  $a_D q_1$  was mandatory in addition to (or regardless of  $g_D p_6$ ), then the advice in Fig. 6.25c is misguided, for it contradicts  $a_D q_1$ . If  $a_D q_2$  is mandatory, then *only* the advice in Fig. 6.26b gives a candidate solution. While mandatory relations do serve for comparison, they do so by strictly discriminating between advice which will give candidate solutions, and any other advice. This is not the case of the optional relation. If  $a_D q_2$  was optional, then only the advice in Fig. 6.26b would include enough information to have a picosolution for  $a_D q_2$ , and while the advice in Fig. 6.25c and Fig. 6.26c does not have enough information for a picosolution to  $a_D q_2$ , the advice in both of these latter cases still gives candidate solutions. This as we said is obviously not the case if  $a_D q_2$  was mandatory, instead of optional. The optional relation then, along with the confidence and preference relations, serves for comparison, but do not discriminate between advice which gives candidate solutions, and advice which fails to do so.

When there are several criteria for comparison in an a-net, and there are two or more candidate solutions, the comparison criteria need not be clear at all on which of the alternative advice to give. Some criteria might favor different candidate solutions than do other criteria. If we add to the a-net in Fig. 6.26c the advisor's preference  $e_A^P(g_A p_1, a_A q_3)$  for  $g_A p_1$  over  $a_A q_3$ , then the advice shown in that figure does include the more likely  $a_D q_1$ , but does not include the preferred  $g_A p_1$ . The a-net itself does not resolve this problem for the advisor, i.e., does not recommend if the advisor should favor advice more likely to be accepted or that which he prefers. This presence of unresolved tradeoffs is a matter that should be resolved by the decision rule, one which would rank the candidate solutions, to single out that which best satisfies specific the relative importance given to the comparison criteria. As that the full range and kind of criteria is not known for every advisor's problem, the general formulation of the advisor's problem needs to remain neutral with regards to the decision rule.

**Definition 6.39. Advisor's problem:****Given:**

- the cues about the decision-maker, i.e., a set of assumptions  $A_D$ , goals  $G_D$ , evaluations  $E_D$ , and commitments  $C_D$ ,
- as well as the advisor's assumptions  $A_A$ , goals  $G_A$ , evaluations  $E_A$ , and commitments  $C_A$ ,

**find** advice  $\mathbf{c}_A^A(A^* \cup G^* \cup C^*)$  such that:

1.  $A^* \subseteq A_D \cup A_A$ ,
2.  $G^* \subseteq G_D \cup G_A$ ,
3.  $C^* \subseteq C_D \cup C_A$ ,
4.  $A^*, G^*, C^* \not\vdash \perp$ ,
5. for every  $pl_i \in \mathbf{M}$ :  $A^*, G^*, C^* \vdash pl_i$ , where  $\mathbf{M}$  is the set of mandatory labeled propositions.

The advisor's problem can also be simply read as: given the cues about the decision-maker and the advisor's assumptions, goals, evaluations, and commitments, find a candidate solution.<sup>8</sup>

Every candidate solution to the advisor's problem must be conflict-free and it must satisfy whatever was marked as mandatory. That everything mandatory must be satisfied seems fairly straightforward, in that the mandatory relation would make no sense if it were otherwise – it is called *mandatory* after all. To say that a candidate solution *satisfies* something is not very careful, since a candidate solution satisfies nothing really. Rather, it has logical consequences, so that here, “satisfies” abbreviates “has as a consequence,” or perhaps the more telling “leads us to conclude.”

A candidate solution tolerates no contradictions. This is perhaps as contentious a requirement as for it to satisfy whatever is mandatory, that is to say, hardly shocking at all. Why this intolerance? There is something twisted in making contradictory recommendations, say, to wage war on a province and not to wage war on the same province. While this can be said, just as anything can be said, the content of the contradictory recommendations seems of little use, although the speech act that states the contradiction may have some pragmatic role in the given context, to confuse, for example, or to point out that. Objections to the acceptance of contradictions are usually some or all of the following. Anything is a logical consequence of a contradiction – a principle called *ex contradictione quodlibet* – so that if one takes a contradiction for granted, he must also take a lot else (everything actually) and thus too much that same way. Nothing is both true and false, so there is something wrong with contradictions. It is also claimed that contradictions cannot be believed rationally: if one does believe them, he/she is irrational, i.e., consistency is an essential

---

<sup>8</sup> This simplified reading is correct, because the fourth condition in Definition 6.39 makes sure that the advice is consistent, while the fifth condition requires that every mandatory labeled proposition is a consequence of the advice. The fourth and fifth conditions together make sure that the closure of the advice that resolves the advisor's problem is a candidate solution, in line with Definition 6.31.

property of the *being rational* concept. Then, there is the idea that if contradictions were acceptable, no one could be rationally criticized, for criticism requires opposition. Finally, if they were acceptable, then no one could deny anything, for their acceptance of an idea would not mean they reject its opposite. Graham Priest, a philosopher makes the case against all of these, thereby arguing that there is perhaps nothing wrong in accepting *some* contradictions (Priest 1998), but not all of them. For instance, the *ex contradictione quodlibet* is a contentious principle, for it assumes that rational belief is closed under entailment, i.e., that one believes every logical consequence of what he believes; in the terminology of AML, the advisor would never “have” only parts of an a-net, but always the entire a-net. The problem with this is that the ability to compute logical consequences and the memory to hold them are scarce, i.e., the cognitive abilities are limited enough to make it difficult to accept that rational belief includes all logical consequences. Arguments in favor of some contradictions also point out that AML is deficient: when there are contradictory assumptions, goals, or otherwise, and there are reasons to accept the contradiction, then AML will not recognize this, it will never include the sides of the contradiction in the same candidate solution.

Apart from the use of the mandatory relation, there is not a single mention of the other unary and binary relations from AML in the problem statement. The only relations relevant for finding candidate solutions are conjunction, implication, inference, and conflict: this we see from the definition of  $\vdash$  and its role in the advisor's problem. The absence of preference, confidence, and optional relations is not troubling though, as they serve for the *comparison* of candidate solutions.<sup>9</sup> As the problem statement gives no suggestions on how to compare, it includes no decision rule. The models of choice we sketched in Chap. 5 all included decision rules, the purpose of which was to single out one of the alternative courses of action. As we shall see below, the various decision rules from Chap. 5 can be reformulated as decision rules for the advisor's problem.

## 6.5 Solving the Advisor's Problem

The solving effort being oriented, rather obviously toward finding a solution, it cannot stop as soon as the candidates are found. The resolution of the advisor's problem plays out in three acts. Start by constructing the explanation and prediction, i.e., the

---

<sup>9</sup> They *could* (but certainly not *must*) be used in finding candidates, but that would require a different formulation of the advisor's problem, one in which the confidence, preference, and optional relations have an explicit role, or in other words, a different definition of the candidate solution concept. If, for instance, a candidate solution were only that which in addition maximizes preference satisfaction (whatever that means precisely), then clearly we would not be looking for every consistent set which satisfies whatever is mandatory, but only those which in addition include most of the preferred nodes of the a-net. This is a very rough idea obviously, as having the maximal number of preferred nodes in the solution need not automatically mean that the solution is the most preferred, as some preferences may be more important than others.



picotheory, by finding cues and relating them by the various relations available in AML. Find, then, the candidate solutions. Compare the candidates, and pick one out.

The making of explanations and predictions involves the finding of cues and of the relations between them, the recording of both in an a-net. When the cues fail to be clear – when they are vague, ambiguous, not detailed or precise enough, and so on – they are *refined*. Given an unclear assumption or goal  $pl$  in the a-net, to refine it is to find some clearer assumptions and/or goals which together have  $pl$  as their logical consequence. To do so is in fact to look for a picosolution for  $pl$ , i.e., some  $\langle \Gamma, pl \rangle$ , whereby the base  $\Gamma$  of the picosolution includes less ambiguous, more detailed, and/or more precise assumptions and/or goals. When, for example, the steps of a plan toward some goal  $X$  are detailed, that detail, the assumptions about the steps to take, or the intermediary goals to achieve, together with  $pl$  are a refinement. Not every picosolution is the result of refinement, even though the base of a picosolution is in the inference relation with the head of the picosolution, and even as to refine is to set a node as head of a picosolution, then search for the base of that picosolution. If a picosolution  $\langle \Gamma, pl \rangle$  is the result of refinement, then its base  $\Gamma$  must be clearer and more detailed than  $pl$ .

The more general point about this first act in resolution is that the advisor may start with cues and his own assumptions, goals, and otherwise, but that those initially considered are not necessarily satisfactory. Machiavelli aims to be general, has thus the luxury of ignoring the peculiarities of the lands and princes his advice may concern. He may advise to overthrow lesser rulers, or make friends with them, but hardly would his advice be as useful as that which says how to do either of these. The construction of explanations and predictions is in one way or another a problem-solving activity. Hence the concept of a picosolution, and the *solution* part in its name: to solve a problem, it has been repeatedly judged useful to brake it down, and attack such parts one after another – a picosolution is a fragment of a candidate solution, its building block, and is the result of whatever problem-solving technique that may have been applied. It is unimportant what technique it is. Both refinement and planning, as well as, say, the specialization of some  $pl$ , if they result in a set of labeled propositions a consequence of which is  $pl$ , result in a base of a picosolution for  $pl$ . Although the problem-solving techniques may be many and variously named, they do manipulate a limited set of ingredients and thereby are amenable to a general classification. The components are those of a picosolution, and there are not so many basic manipulations. To grow an a-net any one of the following, or a combination thereof may be applied:

- Add a node without relating it to other nodes.
- Add a relation on a node, or between existing nodes.
- *Deduction*: Given assumptions and goals, and knowing the consequence relation, determine their consequences and add each to the a-net.
- *Induction*: Given  $pl_1, \dots, pl_n$ , hypothesize the implication  $\phi \equiv pl_1 \wedge \dots \wedge pl_{n-1} \rightarrow pl_n$ , and consequently add the assumption  $\mathbf{k}\phi$  to the a-net, along with the inference relation, in which  $pl_1, \dots, pl_{n-1}, \mathbf{k}\phi$  are premises and  $pl_n$  is the conclusion.

- *Abduction*: Given  $pl_n, \phi \equiv pl_1 \wedge \dots \wedge pl_{n-1} \rightarrow pl_n$ , and the consequence relation, hypothesize the conjunction  $pl_1 \wedge \dots \wedge pl_{n-1}$ , and thus add  $pl_1, \dots, pl_{n-1}$  to the a-net, as well as the inference relation, in which  $pl_1, \dots, pl_{n-1}, \mathbf{k}\phi$  are premises and  $pl_n$  is the conclusion.

Just how far an a-net should be grown has no answer which both is general and useful. The definition of the candidate solution and the problem statement have nothing to say on this. That silence was intended, to keep both of these notions as bases from which to make variants which add further constraints. One variant, appropriate in cases when the advisor ought not advise goals, but rather actions to take can be called the tactician advisor's problem. Compared to the original problem statement, this one says that the advisor should not recommend goals at all, but only assumptions and commitments. Assumptions he should be suggesting should be also about actions, hence the requirements that the set  $\mathbf{A}_{\text{Act}}^*$  of such assumptions cannot be empty (second condition in Definition 6.40), and that it ought to carry assumptions about the conditions in which actions can be executed, that these conditions can be met, and about the effects of the actions. The tactician should recommend not *why* to do something, but *how* to do it.

**Definition 6.40. Tactician advisor's problem:**

**Given:**

- the cues about the decision-maker, i.e., a set of assumptions  $\mathbf{A}_D$ , goals  $\mathbf{G}_D$ , evaluations  $\mathbf{E}_D$ , and commitments  $\mathbf{C}_D$ ,
- as well as the advisor's assumptions  $\mathbf{A}_A$ , goals  $\mathbf{G}_A$ , evaluations  $\mathbf{E}_A$ , and commitments  $\mathbf{C}_A$ ,

**find** advice  $\mathbf{c}_A^*(\mathbf{A}^* \cup \mathbf{G}^* \cup \mathbf{C}^*)$  such that:

1.  $\mathbf{A}^* \subseteq \mathbf{A}_D \cup \mathbf{A}_A$ ;
2.  $\mathbf{A}^* = \mathbf{A}_{\text{Act}}^* \cup \mathbf{A}_{\text{NAct}}^*$  and  $\mathbf{A}_{\text{Act}}^* \cap \mathbf{A}_{\text{NAct}}^* = \emptyset$ , as well as  $\mathbf{A}_{\text{Act}}^* \neq \emptyset$ , i.e., the set  $\mathbf{A}^*$  of assumptions has two nonintersecting parts  $\mathbf{A}_{\text{Act}}^*$  and  $\mathbf{A}_{\text{NAct}}^*$ ;
3.  $\mathbf{A}_{\text{Act}}^*$  is the set of assumptions-about-actions, i.e., assumptions (1) about conditions that ought to hold in order to perform actions, (2) that these conditions can be met, and (3) about the effects of performing these actions;
4.  $\mathbf{C}^* \subseteq \mathbf{C}_D \cup \mathbf{C}_A$ ;
5.  $\mathbf{A}^*, \mathbf{C}^* \not\vdash \perp$ ;
6. for every  $pl_i \in \mathbf{M}$ :  $\mathbf{A}^*, \mathbf{C}^* \vdash pl_i$ , where  $\mathbf{M}$  is the set of mandatory labeled propositions.

If the problem were that of the tactician, the a-net ought to be grown up to the point at which all of its source nodes are either commitments or assumptions, i.e., members of  $\mathbf{C}^* \cup \mathbf{A}_{\text{Act}}^* \cup \mathbf{A}_{\text{NAct}}^*$ .<sup>10</sup>

<sup>10</sup> A note on terminology: in a directed graph, every node which has no outgoing lines is called a *sink* node. Every node which has no incoming lines is called a *source* node. When the graph is a tree, i.e., it has no cycles, a source node is called a *leaf* node, and a sink node is called a *root* node.

The problem tells us what to look for and consequently gives its own conception of what a candidate solution is. Every variant of the advisor's problem comes with its own notion of a candidate solution. It is not difficult to see that the tactician's problem asks for tactical candidate solutions.

**Definition 6.41. Tactical candidate solution:** A set  $\mathcal{S}$  of picosolutions is a tactical candidate solution if and only if:

1.  $\mathcal{S} = \{\langle \Gamma_i, pl_i \rangle \mid \bigcup_{i=1}^n \Gamma_i \not\vdash \perp\}$ ,
2.  $\mathcal{S}$  contains picosolutions for every mandatory labeled proposition, and
3. every  $pl_j \in \bigcup_{i=1}^n \Gamma_i$  for which there is no picosolution is an assumption-about-actions.

Every variant of the original problem statement is its specialization, so that every candidate solution concept is a specialization of the original candidate solution concept in AML. This being the case, whatever the variant of the candidate solution concept, every instance thereof is both an instance of the original candidate solution concept and a candidate solution to the original statement of the advisor's problem.

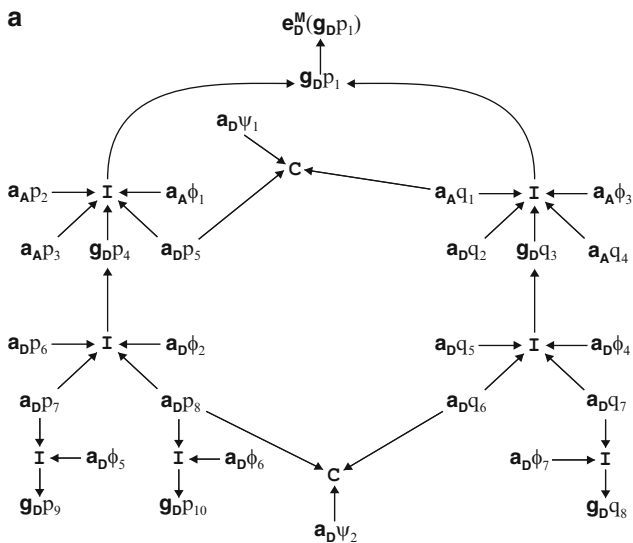
Other specializations of the advisor's problem can be given, without introducing comparison relations. An example are the problem and candidate solution concepts which are counterparts to the tactical variants. A strategist would not be recommending actions to execute, but goals to pursue. The strategist advisor's problem would amount to be to find goals to recommend, not assumptions-about-actions as a tactician would.

Finding all candidate solutions in an a-net is a time-consuming process. It becomes impractical manually for all but the small a-nets of a dozen or two of nodes. The time it takes grows exponentially with the number of nodes.

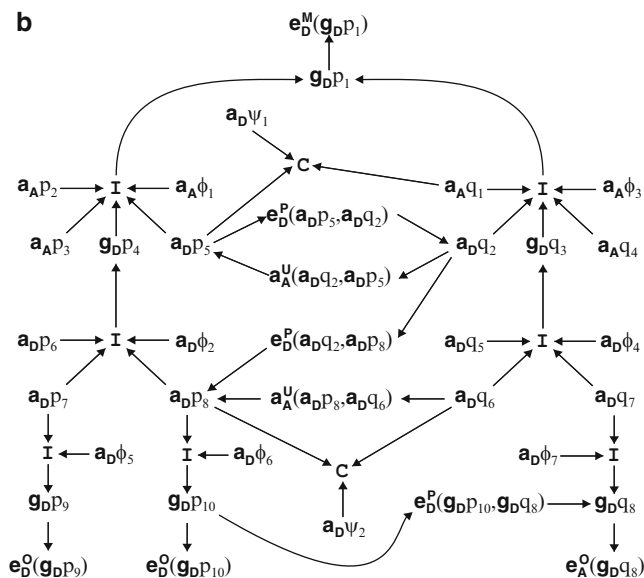
A decision rule sets the requirements that a candidate solution ought to satisfy to be selected over other candidates that have been identified. The preference, optional, and confidence relations serve for the comparison of candidates and the definition of decision rules. Contrast the a-net in Fig. 6.27a and in Fig. 6.27b: the one in Fig. 6.27a contains no information on how to compare the candidate solutions in it, while the a-net in Fig. 6.27b does include mandatory, optional, preference, and confidence relations. A decision rule will say what we ought to conclude from Fig. 6.27b, which of the candidate solutions we ought to select. By adding a decision rule to the original advisor's problem statement, we obtain a specialization of the problem.

Several simple decision rules come to mind:

- MAX-P: Choose the maximal candidate solution which includes the highest number of preferred nodes. Figure 6.28c highlights the candidate solution which satisfies the requirements of this decision rule.
- MAX-O: Choose the maximal candidate solution which includes the highest number of optional nodes. The candidate solution highlighted in Fig. 6.28c has more optional nodes than its counterpart shown in Fig. 6.28d.
- MAX-U: Choose the maximal candidate solution which includes the highest number of confidence nodes. In a confidence relation, the node in which there is more

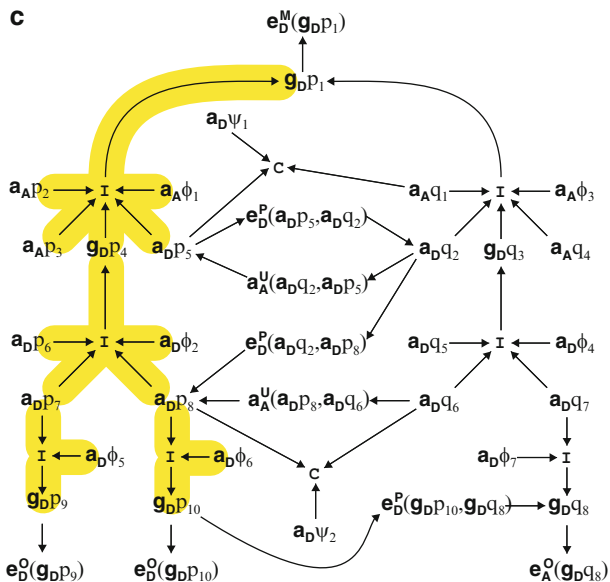


An a-net with two candidate solutions for the goal  $g_D P_1$ . Mandatory, optional, preference, and confidence relations are not shown.

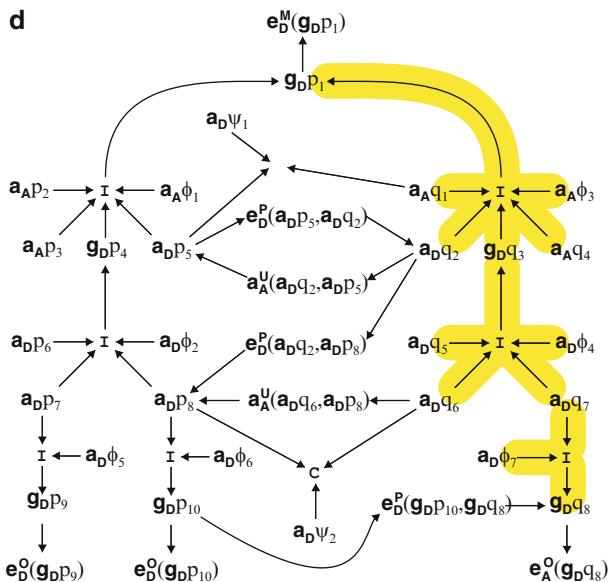


Same a-net as in (a), but some mandatory, optional, preference, and confidence relations are now shown.

**Fig. 6.27** A hypothetical a-net, for which we do not give the readings of the propositions. The variant (a) has no relations that can serve for the comparison of candidate solutions. The variant (b) includes mandatory, optional, preference, and confidence relations allowing thereby the comparison of candidates



The MAX-P decision rule selects the candidate solution, the closure of which is highlighted above. The MAX-O rule also picks out the same candidate solution.



The MAX-U rule selects the candidate solution, the closure of which is highlighted above.

**Fig. 6.28** (Continued from Fig. 6.27.) A hypothetical a-net, same as in Fig. 6.27, where we highlight candidate solutions picked out by the various simple decision rules discussed in the text

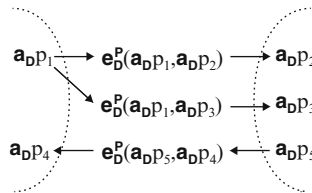
confidence is called the confidence node. Figure 6.28d highlights a candidate solution which satisfies the requirements of the MAX-U rule.

- MAX-PO: Choose the maximal candidate solution which includes both the highest number of preferred and the highest number of optional nodes. Figure 6.28c highlights such a candidate solution.
- MAX-POU: Choose the maximal candidate solution which includes the highest number of preferred nodes, the highest number of optional nodes, and the highest number of confidence nodes. The two candidates, in Fig. 6.28c, d, are in this respect equal, so none of the two is clearly more interesting than the other according to the MAX-POU rule.

Each of these rules is deficient in one way or another. MAX-P disregards entirely optional and confidence relations, MAX-O the preference and confidence relations, and MAX-U the preference and optional relations. Any pairwise combination of MAX-P, and MAX-O, MAX-U will disregard all relations of the kind it does not explicitly consider. MAX-PO will not use the information in confidence relations. Combining all three avoids neglect. While MAX-POU does characterize very attractive solutions, there is obviously no guarantee that every a-net will include at least one such candidate solution. We can respond in two ways to the absence of a suitable candidate. One is to use another decision rule. The other is to revise the a-net to make the candidate which satisfies MAX-POU.

Decision rules that count preferred, optional, and/or confidence nodes are both simple and simplistic. By counting only the preferred nodes, nodes that participate in different numbers of preference relations are assumed equal. If the dashed lines in Fig. 6.29 separate two candidate solutions and we take only the three preference relations for comparison, then the two solutions give us the same count of preferred nodes; each has one preferred node. It is nevertheless obvious that the left-hand side candidate has a node which is preferred in two preference relations, while the one on the right-hand side is preferred in one preference relation. A decision rule which counts the number of preference relations in which nodes of each candidate are preferred will detect this difference.

All comparison relations can be summarized conveniently in a comparison table. Each preference relation, optional relation, and confidence relation obtains its own column in the table (or row, depending on the orientation of the table), and each



**Fig. 6.29** When only the preferred nodes are counted, it is unimportant if a preferred node participates in one or more preference relations. The number of preferred nodes makes for deficient decision rules

**Table 6.2** Comparison table for the a-net in Fig. 6.27b.  $\mathcal{S}_1$  is for the candidate solution highlighted in Fig. 6.28c,  $\mathcal{S}_2$  for that in Fig. 6.28d. *Black dot* indicates which candidate solution includes the preferred, optional, or confidence node

	$\mathcal{S}_1$	$\mathcal{S}_2$
$e_D^P(a_D p_5, a_D q_2)$	•	
$e_D^P(a_D q_2, a_D p_8)$		•
$e_D^P(g_D p_{10}, g_D q_8)$	•	
$e_D^O(g_D p_9)$	•	
$e_D^O(g_D p_{10})$	•	
$e_A^O(g_D q_8)$		•
$a_A^U(a_D q_2, a_D p_5)$		•
$a_A^U(a_D p_8, a_D q_6)$		•

candidate solution takes up a row (or column, depending again on how the table is oriented). A cell of the table tells us if the candidate solution includes the preferred, optional, or confidence node in the relation in the head of the column/row. Every decision rule says how the contents of the comparison table should be used, that is, what conclusions should be drawn from the comparison table. Table 6.2 is the comparison table for the a-net in Fig. 6.27b.

MAX-P only considers the first three rows in Table 6.2, as all preference relations are there. MAX-O applies to the fourth to sixth rows, and MAX-U to the last two rows.

A comparison table for an a-net lists the inputs to a decision rule, the information which serves as a premise to a decision rule. What conclusions will be drawn depend on the decision rule itself, on the way in which it counts, combines, or otherwise manipulates the comparison table. The comparison table gives no numerical values for the relations it mentions. For instance, it has no quantitative estimates of utility or probability, which was intended and due to the intolerance for substitutes argued for in Chap. 5. An important effect is that neither the comparison table nor the a-net from which it was made indicate how the information on preference, optionality, and confidence ought to be combined: it is not clear at all how, for example, we would define the expected utility of a candidate solution. There are in AML no counterparts to multiplication and addition, which apply when quantitative estimates are available. Each preference, optional, and confidence relation serve as criteria independently of one another.

Criteria for the comparison of candidates can also be defined without mentioning comparison relations. They remain outside of the comparison table, and tell us which candidate solutions are robust, clear, and/or efficient, regardless of the specific tastes of the advisor and of the decision-maker. We define these criteria in the following three sections.

## 6.6 Criteria for a Robust Solution

A thing is robust if it can withstand stresses, pressures, or changes procedure or circumstance. The intuitive idea elaborated in this section is that a candidate solution can be more or less robust, that it can be made more robust, and that candidates can be compared in terms of how robust they are.

The closure of a candidate solution gives us a subnet, a part of the a-net in which we found that candidate. It may turn out that the closure of the candidate is the entire a-net, which occurs only if there are no conflicts at all in the a-net – we disregard such cases here, as robustness can only be evaluated when there are conflicts in the a-net.

If there are conflicts in the a-net, the closure of a candidate will be a subnet of the a-net, which means that we can distinguish the information *inside* the candidate subnet from the information *outside* of it. Nodes and relations inside the subnet are all the picosolutions and comparison relations of the candidate solution. The outside information are nodes and relations which are not logical consequences of the candidate solution. *The robustness of a candidate solution depends on how the inside information conflicts with the outside information.*

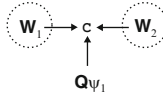
When there is a conflict between an inside node and an outside node, the conflict obviously means that the two are contradictory. The conflict can be read also in another way, compatible with the original interpretation that conflict means contradiction. Namely, the conflict means not only contradiction but also that the outside node is *disputing* the inside node. An example is in the preceding section, in Fig. 6.25. There, the advisor started from the assumptions about how the prince might act to keep the mixed principality without much difficulties and troubles. The advisor then considered why this might not work, namely, that the prince will lose the trust of the lesser rulers in the principality and introduce a mighty foreign ruler, which would result in him losing the principality. The advisor has thereby found arguments against the direction that he believes the prince will follow. The reasons that the advisor, and because of which he judged the prince's choice inappropriate, is information which clearly disputes – or equivalently, counterargues, goes against, questions – the supposed choice of the prince. To make the prince's choice more robust, his decision should include additional information – goals, assumptions, and so on – which disputes the advisor's arguments. Making a candidate solution robust amounts to add more information to it, or otherwise revise it to dispute the outside information which disputes the information inside the candidate solution. We will say that such a candidate *defends* itself. If a candidate solution defends itself, it protects itself from the disputing information outside of it: we can thus say that it is robust with regards to the information it defends itself from.

Figure 6.30 illustrates the role of the conflict relation with regards to the notion of robustness, and in terms of how a candidate solution can be made more robust. The first row of the figure shows the steps we took to add conflicts to the initial one-node a-net, which included only  $W_1$ : we first added the node  $W_2$  and the conflict between  $W_1$  and  $W_2$ , then  $W_3$  and the conflict between  $W_2$  and  $W_3$ , and finally, the node  $W_4$  and the conflict between  $W_3$  and  $W_4$ . The second row reads the conflict relation

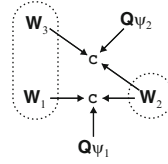


**a**

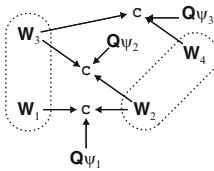
We start from a simple a-net which has only one node. The node is a candidate solution.  $W_1$  follows the notational conventions from Figure 5.23.

**b**

We add information which disputes  $W_1$  and is thereby in conflict with it. There are two candidate solutions now.  $\psi \equiv W_1 \wedge W_2 \rightarrow \perp$ .

**c**

To make the candidate with  $W_1$  more robust, we add to it information  $W_3$  which conflicts with its attacker  $W_2$ .

**d**

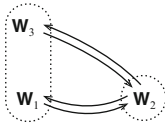
To defend the solution with  $W_2$  from the solution with  $W_1$  and  $W_3$ , we add to it  $W_4$  which conflicts with  $W_3$ .

**e**

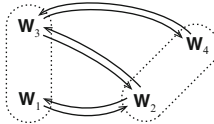
As in (a) above, we have only one node in the a-net. The node is also a picosolution for  $W_1$ , as well as a candidate solution.

**f**

A conflict defined as logical inconsistency, as in AML, has no predefined direction, so we can redraw the conflict in (b) above as two arrows, one from  $W_1$  to  $W_2$ , and another from  $W_2$  to  $W_1$ .

**g**

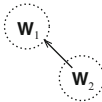
The conflict relations in (c) above are redrawn by following the rule in (f).

**h**

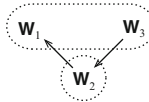
The conflict relations in (d) above are redrawn by following the rule in (f).

**i**

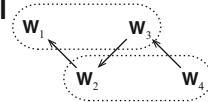
As in (a) and in (e) above, we start from the node  $W_1$ , and the aim is to make more robust the candidate solution which includes that node.

**j**

Since we are interested in making more robust the candidate with  $W_1$ , we only keep one direction in the conflict, indicating thereby an attack on  $W_1$ .

**k**

The conflict relations in (c) above are redrawn by following the rule in (f).

**l**

The conflict relations in (d) above are redrawn by following the rule in (f).

**Fig. 6.30** Role of the conflict relation in making a picosolution and candidate solution robust. The first row of the figure – parts (a)–(d) – show conflict relations being added to a simple one-node a-net. The second row – parts (e)–(h) – redraws the conflict relations from the first row in terms of pairs of arrows, to indicate that in a conflict, nodes attack each other. The third row – parts (i)–(l) – focuses only on some directions of the conflicts, to illustrate the process of making more robust the candidate solution which includes  $W_1$ .

as symmetrical attack: e.g., in Fig. 6.30a, we say that  $\mathbf{W}_1$  is in conflict with  $\mathbf{W}_2$ , but that  $\mathbf{W}_1$  attacks  $\mathbf{W}_2$ , and that  $\mathbf{W}_2$  attacks  $\mathbf{W}_1$ . In the third row, our aim is to improve the robustness of the candidate solution in Fig. 6.30i, so we initially focus on attacks which are directed toward  $\mathbf{W}_1$ : in Fig. 6.30j, we still have the conflict between  $\mathbf{W}_1$  and  $\mathbf{W}_2$ , but since it is the robustness of  $\mathbf{W}_1$  we are evaluating and trying to improve, we only draw the attack from  $\mathbf{W}_2$  to  $\mathbf{W}_1$ , and leave out the one from  $\mathbf{W}_1$  to  $\mathbf{W}_2$ . The attack from  $\mathbf{W}_2$  to  $\mathbf{W}_1$  tells us that we can make the candidate with  $\mathbf{W}_1$  robust only if we find information which is consistent with  $\mathbf{W}_1$  and which attacks  $\mathbf{W}_2$ . This being  $\mathbf{W}_3$ , we add it in Fig. 6.30k and draw the attack from  $\mathbf{W}_3$  to  $\mathbf{W}_2$ . Finally, in Fig. 6.30l, we added  $\mathbf{W}_4$  which makes the candidate with  $\mathbf{W}_2$  robust with regards to  $\mathbf{W}_3$ , since  $\mathbf{W}_4$  attacks  $\mathbf{W}_3$ . *If robustness were the only criterion for the selection of a candidate solution*, then there would be no clear winner between the two candidates in Fig. 6.30l. If that were the only criterion for selecting a candidate in Fig. 6.30k, then we would pick out the candidate with  $\mathbf{W}_1$  and  $\mathbf{W}_3$ , over the one with  $\mathbf{W}_2$ .

In the terminology of attacks and defenses, we can say the following about the third row of Figure 6.30. The candidate in Fig. 6.30j is attacked by  $\mathbf{W}_1$  and does not defend itself from it.  $\mathbf{W}_3$  defends it from  $\mathbf{W}_2$  in Fig. 6.30k. But in Fig. 6.30l, the candidate with  $\mathbf{W}_1$  is attacked by  $\mathbf{W}_4$ , and does not defend itself from it. In Fig. 6.30l, the candidate is robust with regards to  $\mathbf{W}_2$ , but not with regards to  $\mathbf{W}_4$ .

To evaluate the robustness of a candidate solution, we apply the *justification* process. To justify is “to show or maintain the justice or reasonableness of (an action, claim, etc.); to adduce adequate grounds for; to defend as right or proper.” The *Oxford English Dictionary* also tells us that a justification is “the action of justifying or showing something to be just, right, or proper; vindication of oneself or another; exculpation; verification, proof.”

Justification is a confrontation of arguments. An argument can be defined recursively as follows:

1. Any information of the form  $P$  therefore  $c$  is an argument, where  $c$  is called “conclusion” and  $P$  is a set of premises, whereby there is commitment to the truth of the premises.
2. For  $A$  and  $B$  such that  $A = P_A$  therefore  $c_A$  and  $B = P_B$  therefore  $c_B$ , if  $P_A \subseteq P_B$  then  $A$  is a subargument of  $B$ .
3. The conclusion cannot be used to support its premise.
4. Premises must be consistent.
5. Nothing is an argument unless it obeys the rules above.

The suggested conception of an argument is common in philosophy and artificial intelligence.<sup>11</sup> It allows complex arguments, in which a premise can be a conclusion of another argument – as in the second requirement above. It bans cyclical

<sup>11</sup> The definition follows David Hitchcock’s synthesis on the definitions of the concept of argument (Hitchcock 2006). Essentially the same conceptions are used in artificial intelligence research on nonmonotonic reasoning via argumentation (Chesñevar et al. 2000), which emphasize internal consistency, the premises-conclusion structure, and the intervention of an argument in another argument.

arguments via the third requirement. The fourth requirement makes sure that the argument is not internally inconsistent.

An argument in AML can be any internally consistent set of nodes, hence a picosolution or a consistent set of picosolutions. To evaluate the robustness of a candidate solution, we identify all arguments in the candidate which are attacked. Each such argument is analysed separately. For each argument, we identify all arguments which attack it. Then, we look for all arguments (in the candidate we are evaluating) which are defending the attacked argument. As there can be a chain of attacks – i.e., the argument which defends another may be attacked itself, its attacker may be attacked, and so on – the justification process is recursive.

**Definition 6.42. Argument in AML:** An argument is any set

$$\mathcal{A} = \{\langle \Gamma_i, pl_i \rangle \mid 1 \leq i \leq n, n \geq 1, \bigcup_{i=1}^n \Gamma_i \not\vdash \perp\},$$

where  $\langle \Gamma_1, pl_1 \rangle, \dots, \langle \Gamma_n, pl_n \rangle$  are  $n$  picosolutions which are not inconsistent, i.e., their bases are not inconsistent.

A picosolution clearly resembles an argument: it has premises, a conclusion, it is internally consistent. A consistent set of picosolutions can be viewed as an argument for any of the nodes in its closure. An argument can thus in an a-net be a picosolution, a consistent set of picosolutions, a nonmaximal candidate solution, a consistent set of nonmaximal candidate solutions, or a maximal candidate solution. Being defined entirely from concepts available in AML, the argument concept is not a primitive notion in AML.

We saw in Fig. 6.30 that it is useful to read conflict as a symmetrical attack. This understanding of conflict is straightforward when there are only two nodes in a conflict. Figure 6.31 illustrates how conflict is read as attacks when it involves more than two nodes.

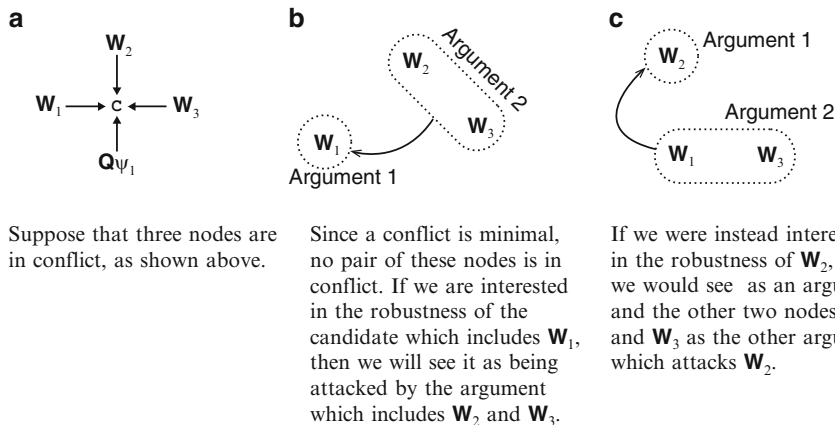
The attack relation is defined between arguments.

*Remark 6.3.* Given an argument  $\mathcal{A} = \{\langle \Gamma_1, pl_1 \rangle, \dots, \langle \Gamma_n, pl_n \rangle\}$ , we denote  $B(\mathcal{A})$  the union of the bases of its picosolutions, i.e.,  $B(\mathcal{A}) \equiv \Gamma_1 \cup \dots \cup \Gamma_n$ .

**Definition 6.43. Attack relation:** An argument  $\mathcal{A}_1$  attacks another argument  $\mathcal{A}_2$  if and only if  $B(\mathcal{A}_1) \cup B(\mathcal{A}_2) \vdash \perp$ .

It is not difficult to see that two arguments will attack each other if and only if there is at least one conflict between the picosolutions that they contain. The definition of attack makes sure that there will be two attack relations between any two arguments which have inconsistent bases. As we said above, we do not always draw all of the attack relations, but instead only those which are relevant to evaluate the robustness of an argument.

The concept of argument and the attack relations are defined from the primitives of AML. This lets us make an *abstract argumentation framework* from every a-net.



Suppose that three nodes are in conflict, as shown above.

Since a conflict is minimal, no pair of these nodes is in conflict. If we are interested in the robustness of the candidate which includes  $w_1$ , then we will see it as being attacked by the argument which includes  $w_2$  and  $w_3$ .

If we were instead interested in the robustness of  $w_2$ , then we would see as an argument and the other two nodes  $w_1$  and  $w_3$  as the other argument which attacks  $w_2$ .

**Fig. 6.31** The way to read n-ary conflicts as attacks depends on which node we want to evaluate for robustness

An abstract argumentation framework is a set of arguments connected by attack relations. The emphasis in it is obviously on conflicts, since only attack relations are available, and thereby not on preference, confidence, mandatory, or optional relations.

Phan Minh Dung, a computer scientist, suggested the concept of an abstract argumentation framework (Dung 1995), as a model of the mechanism that humans use in argumentation and following many prior efforts to arrive at a general and formal model of argumentation. The framework is *abstract* because it considers both the concept of argument and the attack relation as primitives. It leaves them undefined. We can consequently use our own argument concept and our attack relation, and still make a Dung's argumentation framework out of these. We can thus view a-nets as concrete argumentation frameworks, concrete in the sense that they have their own conception of argument and attack. By making an argumentation framework from an a-net, we are taking another perspective on the conflicts in that a-net: the arguments in the a-net are the arguments for or against potential decisions and/or advice. A decision or an advice in the a-net are arguments themselves, as long as they are internally consistent.

That we can make an argumentation framework from an a-net means that we can evaluate the robustness of a candidate solution, or of a picosolution, or of any consistent set of picosolutions by analyzing the argumentation framework. Hence the interest in knowing what an abstract argumentation framework is, as far as the analysis of advice goes.

**Definition 6.44. Argumentation framework:** An argumentation framework is a pair  $(AR, att)$ , where  $AR$  is a set of arguments and  $att \subseteq AR \times AR$  is the attack relation between arguments.

When an argumentation framework is made from an a-net, every argument in  $AR$  is an instance of the argument concept from Definition 6.42 and every attack relation is an instance of the attack relation from Definition 6.43.

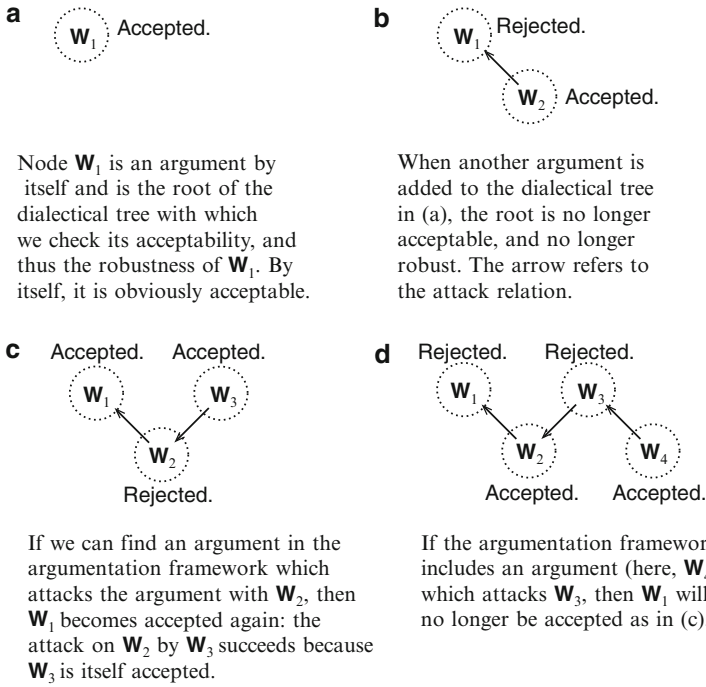
The key question in an argumentation framework is which arguments are acceptable. If we are interested in a single argument, then we rather ask if that argument alone is acceptable. If that argument is actually a candidate solution from an a-net, and if we can show it to be acceptable in the argumentation framework made from an a-net, then that candidate is robust with regards to the entire a-net: no information that attacks it is itself acceptable, or equivalently, no attack on it succeeds. To determine if an argument  $\mathcal{A}$  is acceptable in an argumentation framework  $(AR, att)$ , we apply the justification process.

**Definition 6.45. Justification process:** The justification of an argument  $\mathcal{A}$  in an argumentation framework  $(AR, att)$  consists of recursively defining and labeling a *dialectical tree*  $T(\mathcal{A})$  as follows:

1. The argument  $\mathcal{A}$  is the root of the dialectical tree  $T(\mathcal{A})$ .
2. Every argument  $\mathcal{A}_i \in AR$  which attacks  $\mathcal{A}$  (i.e.,  $(\mathcal{A}_i, \mathcal{A}) \in att$ ) is set as a node of  $T(\mathcal{A})$  and a directed line is drawn from every  $\mathcal{A}_i$  to  $\mathcal{A}$ . Every argument  $\mathcal{A}_i$  is the root of its own dialectical tree  $T(\mathcal{A}_i)$ .
3. When no argument in  $AR$  can be added to  $T(\mathcal{A})$ , label the leaves of the tree **accepted**. For any inner (i.e., non-leaf and non-root) node, label it **undefeated** if and only if every child of that node is a **rejected** node. An inner node will be labeled as **rejected** if and only if it has at least one undefeated node as a child. Label the entire dialectical tree in this way.
4.  $\mathcal{A}$  is acceptable if and only if the node  $\mathcal{A}$  in  $T(\mathcal{A})$  is labeled **undefeated**, and consequently considered acceptable.

Justification is recursive because every attacker of the root node becomes the root of its own dialectical tree in the second step of the justification process. Figure 6.32 illustrates the building and labeling of the dialectical tree in a simple argumentation framework.

A candidate solution can be, as we said earlier, *more or less* robust. Justification is a yes or no affair, in that it says either that an argument is undefeated and consequently acceptable, or the opposite, that the argument is not acceptable. The way in which degrees come into play is via the restriction of the content of the set of arguments  $AR$  in the argumentation framework. If we wish to be very strict, and are interested in arguments which are only acceptable with regards to the entire a-net, then  $AR$  must include all arguments from the a-net. If the argument is acceptable in that case, it cannot be more robust since no attack on it is successful in the a-net. However, if we wish to be lenient, then we would put only some arguments from the a-net in  $AR$ . Suppose that we have two argumentation frameworks,  $(AR, att)$  and  $(AR', att')$ , where  $AR' \subset AR$  and  $att' \subset att$ , so that  $(AR', att')$  is included in and thereby smaller than  $(AR, att)$ . If an argument  $\mathcal{A}_1$  is acceptable with regards to  $(AR', att')$  and  $\mathcal{A}_2$  is acceptable with regards to  $(AR, att)$ , then  $\mathcal{A}_2$  is more robust than  $\mathcal{A}_1$ . It is in this sense that arguments, and thereby candidate solutions, picosolutions, and so on, can be more or less robust.



**Fig. 6.32** An illustration of the justification process. The figure clearly shows how labels alternate on the root node in its dialectical tree, and this depending on the additional arguments and their attack relations. Note that any set of arguments in which no argument attacks other arguments, i.e., an internally consistent set of arguments, can itself be grouped into an argument: above,  $W_1$  and  $W_3$  can be in the same argument, or in the same candidate solution, just as  $W_2$  and  $W_4$  can; however,  $W_1$  and  $W_2$  cannot go together into an argument, and thereby into a candidate solution

The more of a candidate solution are justified arguments, the more robust that candidate solution is. If we disregard all other criteria for the comparison of candidate solutions, and commit to pick the most robust one, then we are committing to a particular conception of rationality: we assume that the advice, the candidate solution which we choose to communicate, will be accepted because the arguments in it are justified. We would thereby assume that the decision-maker decides whether something is believable through argumentational reasoning, as Dung calls it: “the idea of argumentational reasoning is that a statement is believable if it can be argued successfully against attacking arguments. [...] whether or not a rational agent believes in a statement depends on whether or not the argument supporting this statement can be successfully defended against counterarguments.” (Dung 1995, p. 323) Very roughly speaking, this sort of rational agent would decide what is believable by disregarding preference, confidence, and other comparison relations. Regardless of how this conception of rationality may be off, it is clear that justifying an argument leads us to consider the information that counters it in the a-net. This alone is useful enough, for it leads us to consider how we could defend the argument in the a-net.

Making sure that a candidate solution is robust is a valid aim, but not if taken alone. If we idiotically disregard other criteria, it will actually be trivial to make any candidate solution robust. The formalization of argumentation above cannot distinguish relevant from irrelevant arguments, it cannot perform the interpretation we discussed in Chaps. 3 and 4. The dialectical tree can be in a sense manipulated then, by adding irrelevant arguments which will neutralize the attacks on the root node. As described here, argumentation is a relevant toolset to make candidates robust, but only within a broader approach which considers the interpretation in context of the a-net, as well as the criteria we discuss in the next two sections.

## 6.7 Criteria for a Clear Solution

Lack of clarity will manifest itself as polysemy or vagueness. The former designates the openness of reference, the possibility for one among multiple potential referents to be intended. The latter involves uncertainty as to whether a property applies, how big a thing ought to be to be big, when it is in contrast small, beautiful or ugly, tall or short.

It is context, the information that it provides and in which communication happens that might be good enough to resolve quarrels of intended reference in cases of polysemy. Additional detail may be given to clarify, that is, a piece of advice may be refined – instead of recommending only the conclusion of one's thinking, laying out the premises would clarify. Vagueness in contrast might involve added detail of a particular kind, a standard of comparison, the scale and the cutoff points.

Clarity or lack thereof in a candidate solution is to be evaluated through considerations that we already outlined and debated in Chaps. 3 and 4. Criteria for evaluating the clarity of a candidate solution are, then, the characteristics of its reference relation.

An obvious and general rule is accessible, seems uncontroversial enough, and might even be thought of as a good style of advising, just as some have argued it to be good style in art: "Bad style, as critics discuss it, might then be tentatively described in these terms: the diction and syntax of a discourse are such as to produce [...] ambiguity and obscurity." (Beardsley 1958, p. 227).

The injunction to be clear, to pinpoint the reference with precision, is only as useful as it serves to the purpose the advisor chooses for his recommendations. Whether the syntax seems to refer to a third party to proper referents, and do so well enough, is irrelevant really, as long as it is tailored to be interpreted in the desired way by the decision-maker. When the Archbishop of Canterbury tells Henry the Fifth in Shakespeare's *Henry V* (quoted earlier) that "In aid whereof we of the spirituality / Will raise your highness such a mighty sum / As never did the clergy at one tome / Bring to any of your ancestors," we might expect Henry to ask "how much exactly?". If he does not, and still chooses war over peace, then the Bishop's suggestion was clear enough. Although it is vague, its purpose is to support the decision to go to war, not to go into how much the clergy could give.

Advice is clear *with regards to its purpose* if it is communicated in such form and way, if its syntax, semantics, and pragmatics are such that they recommend actions toward the realization of that purpose. Whether individual words, sentences, or otherwise are clear independently of that purpose is not necessarily important. When Winston Churchill delivered said “We shall fight on the beaches, we shall fight on the landing grounds, we shall fight in the fields and in the streets, we shall fight in the hills; we shall never surrender” in his speech to the British Parliament on 4 June 1940, he did not go on to list the specific beaches and landing grounds. The syntax and the pragmatics of the speech made this unnecessary, for they fit perfectly the motivational purpose of the address. What does matter is that whatever remains unclear does not hamper the realization of the aim of the advice, if it stands in the way of reaching that aim. Spotting deficiencies of clarity need not therefore be a call to its suppression at all costs, only the suppression of that which stands in the way of the objectives sought when advising.

It was argued that in art “a style may be said to be appropriate to, or coherent with, the rest of the work’s meaning if, so to speak, the eddies of meaning it sets up work together with the main streams.” (Beardsley 1958, p. 227) In advice, the choice of syntax and pragmatics, the style if one wishes to call it so, will be at fault if such choices put advice in the way of the aims that it was intended to promote.

## 6.8 Criteria from Empirical Evidence

Research in psychology can inform choices of robustness and clarity. The anticipated goals and preferences of the decision-maker, especially when they do in fact converge with his behavior, are not pieces of information that can be eliminated by mere twist of argument. Offering information that can be verified, that may have been rigorously acquired may not do either. The actual goals and preferences which motivate reasoning – i.e., the forming of impressions, beliefs, attitudes, the evaluation of evidence and the making of decisions – influence strategies for accessing, constructing, and evaluating beliefs. As Ziva Kunda, a social psychologist observes in a survey of psychological research on motivated reasoning (Kunda 1990), there is considerable empirical evidence that motives do affect reasoning, and thereby decision-making. They will, for instance influence the decision-maker’s involvement in the analysis of information; when there is little interest in the outcome of the reasoning task, less effort will be invested than if interest was more significant. Offering particularly robust advice to a decision-maker uninterested at the decision problem may thus prove useless.

When confronted with information which conflicts with already held beliefs, the decision-maker will want to relieve cognitive dissonance by avoiding or ignoring the challenging information, discrediting its source, or argue substantively against it (DiMaggio 1997; Kunda 1990). There is also evidence that “the cognitive exercise of generating counterarguments often has the ironic effect of solidifying and strengthening the original opinion leading to entrenched, polarized attitudes” (Prasad et al. 2009, pp. 2–3).



**Table 6.3** Findings relevant to the choice of a candidate solution, of its robustness and clarity. Literature that may be relevant is extensive; Daniel Gilbert's and Ziva Kunda's surveys (Gilbert 1991; Kunda 1990) are excellent starting points, and served as guides for findings in the table

---

Findings

---

To verify a proposition, i.e., evaluate if it correctly characterizes its referents, it appears that people (a) mentally represent both the proposition (e.g., "the Eiffel tower is in Paris") and the state of affairs it purports to describe (e.g., a picture of the Eiffel tower) and (b) they compare the two representations and decide whether the proposition is true. The proposition is, however, initially assumed to be true: it is considered false only if discrepancies are found between the two representations. (Clark and Clark 1977)

To comprehend a negated proposition (e.g., "Politician not linked to mafia"), it seems that the person must comprehend the core and positive proposition ("Politician linked to mafia"). Subjects who read denials were left with more negative impressions than those who read neutral assertions. (Wegner et al. 1981)

Invalid information is used in decision-making, even when recognized by as invalid by the decision-maker: in an experiment, subjects were given invalid feedback on their performance; they kept believing that feedback even after the evaluator confessed its falsity. (Ross et al. 1975)

People are not very good at ignoring, forgetting, rejecting, or otherwise failing to believe that which they already comprehended. (Bjork 1972)

When motivated to arrive at accurate conclusions, people invest more cognitive effort, attend to relevant information more carefully, and process it more deeply, often using more complex rules. Accuracy motives do not, however, eliminate by themselves faulty reasoning procedures that the person may apply (e.g., Fischhoff 1977). (McAllister et al. 1979)

Motives may bias the reconstruction of one's past behavior: when led to believe that toothbrushing or caffeine consumption was bad for their health, subjects reported that they had performed those behaviors in recent past less frequently than did subjects led to believe that the same behaviors were good for their health. (Ross et al. 1981)

---

Acceptance or rejection of advice can be influenced by resource depletion, another name for fatigue and/or distraction: "[w]hen resource-depleted persons are exposed to doubtful propositions (i.e., propositions that they normally would disbelieve), their ability to reject those propositions is markedly reduced" (Gilbert 1991, p. 111). Resource-depletion impairs cognitive mechanisms which would result in the rejection of doubtful propositions:

When one disables the assessment mechanism [...] one should, in fact, find a person with a belief in the comprehended proposition. One should find a person who has been artificially reduced to a state of primitive credulity, a state of judgmental innocence in which everything that is, is true. And this is what one finds. (Gilbert 1991, p. 111)

An autobiographical proposition is a person's verbal and/or behavioural description of herself. Autobiographical propositions tend to be accepted (Ross 1977), even when it may be clear that such a proposition is delivered to mislead or, say, because of a pressure from an authority. Initially accepted autobiographical propositions will be rejected only if there is time and energy for their assessment.

Empirical findings can relevantly inform the choice of a candidate solution, the robustness and clarity it should have, as well as of the syntax and pragmatics in which it will be communicated. Such findings act as criteria for the comparison of candidates, independently of the specific confidence, preference, and other relations which arise from attitudes. Some are listed in Table 6.3. In general, any finding which provides insight into how elements of context may influence the formation of reference relations, the reaction to both confirming and contradicting information, and the formation of preferences will prove invaluable in the identification of cues, their relation, and the drawing of conclusions therefrom.



## Chapter 7

# Perspectives

Five arguments were presented and defended in this book. The first argument is that advice is a tool of coordination. Analysis of advice is thereby a way to understand certain aspects of mechanisms of coordination, and in particular to study the characteristics of information that is exchanged as advice within such mechanisms. Design of advice is in this perspective a way to influence and develop mechanisms of coordination. Neither the analysis nor the design of advice, as they have been presented here, can capture the full complexity of coordination: for instance, the role of sanctions was seldom mentioned, and the role of incentives in coordination was not discussed; both sanctions and incentives are relevant for coordination. They have been studied to a considerable extent elsewhere already.

The second argument is that a specialist turns into a generalist who must rely on advice in decision situations outside his own specialty. Advice will consequently affect his decision autonomy. However undesirable this may be, the specialist-turned-generalist cannot avoid the fundamental tradeoff between the autonomy of choice and the necessity to coordinate. The way in which he analyzes and designs advice will influence his autonomy of choice and his position within mechanisms of coordination.

The third argument is that analysis and design of advice require a method which acknowledges that both the production and consumption of advice usually happen in decision situations characterized by variously imprecise, unclear, incomplete, or conflicting information about phenomena that are not understood to the extent that they can be measured toward quantification and numerical analysis. The consequence of this argument was the departure from the tools of classical decision theory and of decision analysis.

The fourth argument acknowledged that advice cannot be properly analyzed and designed without understanding the role that reference relations play, the advantages and disadvantages of multiple reference, the basic notions of signs, concepts, and objects, and more generally the role of communication in the production and consumption of advice. This argument has an additional reading, namely that analysis and design of advice can use the theoretical tools of ontology engineering, as well as of conceptual modeling and analysis.

The final, fifth argument is that the rationale of advice can be analyzed both to decide whether to accept advice and to design advice. Rationale is analyzed through a qualitative, symbolic framework defined as a simple mathematical logic.

A general method for the analysis and design of advice and choice was presented through these arguments. The method relies on a number of assumptions which can and should be challenged, so that this or another method can be developed further. Chapter 6 in particular defined languages for the modeling of advice, which are actually very simple. There could be merits in going toward more elaborate mathematical formalisms for the modeling and analysis of advice. The interest in formalisms is not only anchored in the willingness to try to be explicit and precise, but also in the possibility of automating tasks in the analysis and design of advice, such as, e.g., the search for candidate solutions.

To sum up, this book offers a general qualitative method for the analysis and design of advice and choice. It can also be read as a case study in method engineering, one which started by making the case for the relevance of analyzing and designing advice, defined the basic inputs and outputs of the method through an ontology of choice and advice, and offered guidelines on how to transform such information to help decide whether to accept advice, and how to design it. Finally, the book illustrates how to depart from classical theoretical means for the study of decisions to build new concepts, relations, and formalisms for a qualitative analysis of decision situations, and of the role of advice therein.

# References

- Adams D (1979) *The Hitchhiker's Guide to the Galaxy*. Pan Books
- Adams F, Steadman A (2004) Intentional action in ordinary language: core concept or pragmatic understanding? *Analysis* 64(2):173–181
- Adolphs R (2002) Neural systems for recognizing emotion. *Current Opinion in Neurobiology* 12:169–177
- Akst D (August 4, 2002) On the Contrary: Was Gordon Gekko Right About Greed? *The New York Times*
- Alexander RM (2003) Modelling approaches in biomechanics. *Philosophical Transactions: Biological Sciences* 358(1437):1429–1435
- Anderson RL (1998) Truth and Objectivity in Perspectivism. *Synthese* 115(1):1–32
- Anonymous (2000a) *The American Heritage Dictionary of the English Language*. Houghton Mifflin Company
- Anonymous (2000b) Code of Ethics. Enron Corporation
- Anonymous (2008) Index of Democracy. The Economist Intelligence Unit
- Anonymous (2009) Democratic party platform of 1840. In: Woolley JT, Peters G (eds) *The American Presidency Project*, University of California, <http://www.presidency.ucsb.edu/>
- Anonymous (May 16, 2002) Is greed good? *The Economist*
- Anonymous (May 30, 2005) French say firm “No” to EU treaty. *The British Broadcasting Corporation*
- Apperly IA (2008) Beyond Simulation-Theory and Theory-Theory: Why social cognitive neuroscience should use its own concepts to study “theory of mind”. *Cognition* 107:266–283
- Arrow KJ (1950) A Difficulty in the Concept of Social Welfare. *The Journal of Political Economy* 58(4):328–346
- Augustine (1963) *The Confessions of St. Augustine*. Penguin Books, Rex Warner's translation.
- Aumann RJ, Hart S (2003) Long cheap talk. *Econometrica* 71(6):1619–1660
- Austin JL (1962) *How To Do Things With Words*. Harvard University Press
- van der Auwera J (ed) (1980) *The Semantics of Determiners*. Routledge
- Bach K (1994) *Thought and Reference*. Oxford University Press
- Baldwin R, Widgrén M (2007) Does the EU need a new Treaty? Centre for Economic Policy Research – Policy Insight (3)
- Bandura A (1974) Behavior Theory and Models of Man. *American Psychologist* 29(12):859–869
- Bealer G (1992) The Incoherence of Empiricism. *Proceedings of the Aristotelean Society, Supplementary Volumes* 66:99–143
- Beardsley M (1958) *Aesthetics: Problems in the Philosophy of Criticism*. Harcourt, Bruce & World
- Becker GS, Murphy KM (1992) The division of labor, coordination costs, and knowledge. *The Quarterly Journal of Economics* 107(4):1137–1160
- Belnap N (1993) On Rigorous Definitions. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 72(2/3):115–146

- Benjamin W (1999) *The Work of Art in the Age of Mechanical Reproduction*. In: Arendt H (ed) *Illuminations*, Pimlico
- Bentham J (1823) *An Introduction to the Principles of Morals and Legislation*. W. Pickering
- Bentham J (1825) *The Rationale of Reward*. J. & H. L. Hunt
- Berman HJ (1983) *Law and Revolution: The Formation of the Western Legal Tradition*. Harvard University Press
- Bierce A (1906) *The Devil's Dictionary*. Doubleday, Page & Company
- Binmore K (2009) Interpersonal Comparison of Utility. In: *The Oxford Handbook of Philosophy of Economics*. Oxford University Press, chap 20
- Bizer GY, Barden JC, Petty RE (2003) Attitudes. In: *Encyclopedia of Cognitive Science*, MacMillan
- Bjork RA (1972) Theoretical implications of directed forgetting. In: Melton AW, Martin E (eds) *Coding processes in human memory*, Winston, pp 217–235
- Bobo L (1997) The color line, the Dilemma, and the dream: race relations in America at the close of the twentieth century. In: Higham J (ed) *Civil Rights and Social Wrongs: Black-White Relations Since World War II*, Pennsylvania State University Press, pp 31–55
- Bonaccio S, Dalal RS (2006) Advice taking and decision making: An integrative literature review and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes* 101:127–151
- Bratman M (1987) *Intentions, Plans, and Practical Reason*. Harvard University Press
- Bush GW (Sept. 20, 2001) Speech to the Joint Session of Congress. The Whitehouse
- Bybee JS (2002) Memorandum for Alberto R. Gonzales, Counsel to the President. U.S. Department of Justice, Office of Legal Counsel
- Camerer C, Loewenstein G, Prelec D (2005) Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature* 43(1):9–64
- Carnap R (1956) Empiricism, semantics and ontology. In: *Meaning and Necessity*, University of Chicago Press
- Casati R, Varzi A (2008) Events. In: Zalta EN (ed) *The Stanford Encyclopedia of Philosophy*, fall 2008 edn
- Castelnuovo G (1919) *Calcolo delle probabilità*. Albrighi e Segati
- Catholic Church (1992) *Catechism of the Catholic Church*. URL <http://www.vatican.va/archive/catechism/>
- Chesñevar CI, Maguitman AG, Loui RP (2000) Logical models of argument. *ACM Computing Surveys* 32(4):337–383
- Churchich N (1990) *Marxism and alienation*. Fairleigh Dickinson University Press
- Churchland PM (1981) Eliminative Materialism and the Propositional Attitudes. *The Journal of Philosophy* 78(2):67–90
- Cieply M (May 5, 2007) Film's Wall Street Predator to Make a Comeback. *The New York Times*
- Clark HH, Clark EV (1977) *Psychology and language: An introduction to psycholinguistics*. Harcourt, Brace, Jovanovich
- Clokier HM (1949) The modern party state. *The Canadian Journal of Economics and Political Science* 15(2):139–157
- Coase RH (1937) The nature of the firm. *Economica* 4(16):386–405
- Coase RH (1988) *The Firm, the Market and the Law*. In: *The Firm, the Market and Law*, University of Chicago Press
- Cohen MD, March JG, Olsen JP (1972) A garbage can model of organizational choice. *Administrative Science Quarterly* 17(1):1–25
- Collins FS, Morgan M, Patrinos A (2003) The human genome project: Lessons from large-scale biology. *Science* 300:286–290
- Cooke RM (2004) Conceptual fallacies in subjective probability. *Topoi* 5(1):21–27
- Crane T (2001) *Elements of Mind: An Introduction to the Philosophy of Mind*. Oxford University Press
- Crespo R (2008) 'The Economic' According to Aristotle: Ethical, Political and Epistemological Implications. *Foundations of Science* 13:281–294

- Crockett H (2001) The Incendiary Pamphlet: David Walker's Appeal in Georgia. *The Journal of Negro History* 86(3):305–318
- Cronin TE (1989) *Direct Democracy: The Politics of Initiative, Referendum and Recall*. Harvard University Press
- Dahl RA (1971) *Polyarchy: Participation and Opposition*. Yale University Press
- Damasio AR (1994) *Descartes's Error: Emotion, Reason and the Human Brain*. Avon
- Damasio AR, Tranel D, Damasio H (1991) Somatic markers and the guidance of behavior: theory and preliminary testing. In: Levin HS, Eisenberg HM, Benton HM (eds) *Frontal Lobe Function and Dysfunction*, Oxford University Press, pp 217–229
- Darnton R (1995) Censorship, a Comparative View: France, 1789 – East Germany, 1989. *Representations* 49:40–60
- Davidson D (1973–1974) On the Very Idea of a Conceptual Scheme. *Proceedings and Addresses of the American Philosophical Association* 47:5–20
- Dennett DC (1978) *Brainstorms*. MIT Press
- Dennett DC (1990) True believers: The intentional strategy and why it works. In: Lycan WG (ed) *Mind and Cognition: A Reader*, Blackwell, pp 150–167
- DiMaggio P (1997) Culture and Cognition. *Annual Review of Sociology* 23:263–287
- Donoghue D (2003) “Moby-Dick” after September 11th. *Law and Literature* 15(2):161–188
- Druckman JN (2001) Using credible advice to overcome framing effects. *Journal of Law, Economics, and Organization* 17:62–82
- Dung PM (1995) On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence* 77(2):321–358
- Dunn BD, Dalgleish T, Lawrence AD (2006) The somatic marker hypothesis: A critical evaluation. *Neuroscience and Biobehavioral Reviews* 30:239–271
- Eagly A, Chaiken S (1996) Attitude structure and function. In: *The Handbook of Social Psychology* (4th ed.), McGraw-Hill
- Eco U (1976) *A Theory of Semiotics*. Indiana University Press
- Eco U (1987) Meaning and Denotation. *Synthese* 73(3):549–568
- Eco U (1999) *Kant and the Platypus: Essays on Language and Cognition*. Harvest Books
- Elster J (1998) Emotions and economic theory. *Journal of Economic Literature* 36(1):47–74
- Feng B, MacGeorge EL (2006) Predicting receptiveness to advice: Characteristics of the problem, the advice-giver, and the recipient. *Southern Communication Journal* 71(1):67–85
- de Finetti B (1974) *Theory of Probability*, vol 1. Wiley
- Fischhoff B (1977) Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance* 3:349–358
- Fodor JA (1975) *The Language of Thought*. Cromwell
- Frege G (1879) *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. L. Nebert
- Frege G, Geach PT, Black M (1951) On concept and object. *Mind* 60(238):168–180
- Friedman M (2002) *Capitalism and Freedom*. Chicago University Press, 1st edn. in 1962
- Frith CD, Frith U (2006) The Neural Basis of Mentalizing. *Neuron* 50:531–534
- Gallese V, Keysers C, Rizzolatti G (2004) A unifying view of the basis of social cognition. *Trends in Cognitive Sciences* 8(9):396–403
- Gangemi A, Guarino N, Masolo C, Oltramari A (2001) Understanding top-level ontological distinctions. In: *International Joint Conference on Artificial Intelligence (IJCAI) Workshop on Ontologies and Information Sharing*
- Gärdenfors P (1996) Mental Representation, Conceptual Spaces and Metaphors. *Synthese* 106(1):21–47
- Gardner PH, Berry DC (1995) The effect of different forms of advice on the control of a simulated complex system. *Applied Cognitive Psychology* 9(7):S55–S79
- Gauker C (2008) Zero tolerance for pragmatics. *Synthese* 165(3):359–371
- Gefer P (July 23, 2009) Icons as Facts, Fiction and Metaphor. URL <http://lens.blogs.nytimes.com/2009/07/23/essay-4/>
- Gilbert DT (1991) How Mental Systems Believe. *American Psychologist* 46:107–119



- Gino F (2006) Do we listen to advice just because we paid for it? The impact of cost of advice on its use. *Harvard Business School*
- Goldsmith DJ, Fitch K (1997) The normative context of advice as social support. *Human Communication Research* 23(4):454–476
- Guarino N, Welty C (2002) Evaluating Ontological Decisions with OntoClean. *Communications of the ACM* 45(2):61–65
- Hagoort P, Hald L, Bastiaansen M, Petersson KM (2004) Integration of Word Meaning and World Knowledge in Language Comprehension. *Science* 304(5669):438–441
- Harries C, Yaniv I, Harvey N (2004) Combining advice: the weight of a dissenting opinion in the consensus. *Journal of Behavioral Decision Making* 17(5):333–348
- Harvey N, Fischer I (1997) Taking advice: accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes* 70(2):117–133
- Hayek FA (2001) *The Road to Serfdom*. Routledge, 1st edn. in 1944
- Hazewinkel M (ed) (2002) *Encyclopedia of Mathematics*. Springer-Verlag
- Heath C, Gonzalez R (1995) Interaction with others increases decision confidence but not decision quality: evidence against information collection views of interactive decision-making. *Organizational Behavior and Human Decision Processes* 61(3):305–326
- Heath J (2009) Methodological individualism. In: Zalta EN (ed) *The Stanford Encyclopedia of Philosophy*, summer 2009 edn
- Heather PJ (1948) Colour Symbolism: Part I. *Folklore* 59(4):165–183
- Hedges C, Al-Arian L (July 30, 2007) The Other Iraq War: Iraq Vets Bear Witness. *The Nation*
- Heider F, Simmel M (1944) An experimental study of apparent behavior. *The American Journal of Psychology* 57(2):243–259
- Herman ES, Chomsky N (1988) Manufacturing Consent: A Propaganda Model. Pantheon
- Hitchcock D (2006) The concept of argument, and informal logic. In: J Gabbay JW P Thagard (ed) *Philosophy of Logic, Handbook of the Philosophy of Science 5*, Elsevier
- Hjek A (2009) Interpretations of probability. In: Zalta EN (ed) *The Stanford Encyclopedia of Philosophy*, spring 2009 edn
- Hoeschele W (2002) The Wealth of Nations at the Turn of the Millennium: A Classification System Based on the International Division of Labor. *Economic Geography* 78(2):221–244
- Hofweber T (2005) A Puzzle about Ontology. *Noûs* 39(2):256–283
- Hollander P (1988) *The Many Faces of Socialism: Comparative Sociology and Politics*. Transaction Publishers
- Howard RA (1966) Decision analysis: Applied decision theory. In: *Proceedings of the 4th International Conference on Operational Research*
- Howard RA (1980) An Assessment of Decision Analysis. *Operations Research* 28(1):4–27
- Hsee CK (1996) The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes* 67(3):247–257
- Hume D (1739–1740) *A Treatise of Human Nature: Being an Attempt to introduce the experimental Method of Reasoning into Moral Subjects*
- Hyde D (2008) Sorites paradox. In: Zalta EN (ed) *The Stanford Encyclopedia of Philosophy*, fall 2008 edn
- Ingram E (1980) Great Britain's Great Game: An Introduction. *The International History Review* 2(2):160–171
- Jeffrey R (2004) *Subjective Probability: The Real Thing*. Cambridge University Press
- Jennings MM (2003) A primer on Enron: Lessons from a perfect storm of financial reporting, corporate governance and ethical culture failures. *California Western Law Review* 39:163–262
- Joyce IT (1997) The Spiny-Lobster Fishery in Cuba. *Geographical Review* 87(4):484–503
- Juran JM (1951) *Quality Control Handbook*. McGraw-Hill
- Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–291
- Kahneman D, Ritov I, Schkade D (1999) Economic preferences or attitude expressions?: An analysis of dollar responses to public issues. *J Risk and Uncertainty* 19

- Kaufmann D, Kraay A, Mastruzzi M (2008) Governance Matters VII: Aggregate and Individual Governance Indicators 1996–2007. World Bank Policy Research Working Paper 4654
- Keeney RL (1982) Decision analysis: An overview. *Operations Research* 30(5):803–838
- Keeney RL, Raiffa H (1976) *Decisions with Multiple Objectives*. Wiley
- Keller KL (1993) Conceptualizing, Measuring, and Managing Customer-Based Brand Equity. *The Journal of Marketing* 57(1):1–22
- Kennedy C (1999) Gradable Adjectives Denote Measure Functions, Not Partial Functions. *Studies in the Linguistic Sciences* 29(1):66–80
- Kennedy C (2007) Vagueness and grammar: the semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy* 30(1):1–45
- Keys P (1991) Operational Research in Organizations: A Metaphorical Analysis. *The Journal of the Operational Research Society* 42:435–446
- Knapp ML, Hall JA (1978) *Nonverbal Communication in Human Interaction*. Duxbury
- Kolmogorov AN (1956) *Foundations of the Theory of Probability*. Chelsea Publishing Company, 2nd English ed., Nathan Morrison's translation. Original published in 1933
- Kunda Z (1990) The Case for Motivated Reasoning. *Psychological Bulletin* 108(3):480–498
- Kuwada K (1998) Strategic learning: The continuous side of discontinuous strategic change. *Organization Science* 9(6):719–736
- Labib SY (1969) Capitalism in medieval islam. *The Journal of Economic History* 29(1):79–96
- Landau E (1951) *Foundations of Analysis*. Chelsea
- Langley A, Mintzberg H, Pitcher P, Posada E, Saint-Macary J (1995) Opening up decision making: The view from the black stool. *Organization Science* 6(3):260–279
- Langton SRH, Watt RJ, Bruce V (2000) Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences* 4(2):50–59
- Leiserson A (1957) The place of parties in the study of politics. *The American Political Science Review* 51(4):943–954
- Leslie DA (1995) Global Scan: The Globalization of Advertising Agencies, Concepts, and Campaigns. *Economic Geography* 71(4):402–426
- Lévi-Strauss C (1961) *Tristes Tropiques*. Criterion Books, 1st edn. in 1955
- Lévi-Strauss C (1995) *Myth and Meaning: Cracking the Code of Culture*. Schocken
- Lieber F (1863) Instructions for the Government of Armies of the United States in the Field. In: Correspondence, Orders, Reports, and Returns of the Union Authorities from January 1 to December 21, 1863
- Lippmann W (1922) *Public Opinion*. Free Press, 1st edn. in 1922
- Locke J (1696) *The Reasonableness of Christianity, as Delivered in the Scriptures*. Awnsham and John Churchill
- Locke J (1825) *An Essay Concerning Human Understanding*, 25th edn. Samuel Marks
- Loewenstein K (1937) Dictatorship and the German constitution: 1933–1937. *The University of Chicago Law Review* 4(4):537–574
- Loomes G, Sugden R (1982) Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal* 92(368):805–824
- Lotto RB, Purves D (2002) A rationale for the structure of color space. *Trends in Neuroscience* 25(2):84–89
- Machiavelli N (2009) *The Prince*. Oneworld Classics, J.G. Nichol's translation. Original published in 1513
- MacLeod WC (1931) *The Origin and History of Politics*. John Wiley & Sons, Inc
- Marai S (2009) Photograph: An Afghan woman displays her finger marked with indelible ink after casting her vote at a polling station in Kabul on August 20, 2009. Agence France-Presse/Getty Images
- Marchionatti R, Gambino E (1997) Pareto and Political Economy as a Science: Methodological Revolution and Analytical Advances in Economic Theory in the 1890s. *The Journal of Political Economy* 105(6):1322–1348
- Marks JH (2005) Doctors of Interrogation. *The Hastings Center Report* 35(4):17–22

- Masolo C, Borgo S, Gangemi A, Guarino N, Oltamari A, Schneider L (2003) DOLCE: a Descriptive Ontology for Linguistic and Cognitive Engineering. Tech. rep., Institute of Cognitive Science and Technology, Italian National Research Council
- McAllister DW, Mitchell TR, Beach LR (1979) The contingency model for the selection of decision strategies: An empirical test of the effects of significance, accountability, and reversibility. *Organizational Behavior and Human Performance* 24(2):228–244
- McGinn C (1981) The Mechanism of Reference. *Synthese* 49(2):157–186
- McGrath M (2008) Propositions. In: Zalta EN (ed) *The Stanford Encyclopedia of Philosophy*, fall 2008 edn
- Mele A (2001) Acting intentionally: probing folk notions. In: Malle B, Moses L, Baldwin D (eds) *Intentions and Intentionality: Foundations of Social Cognition*, MIT Press, pp 27–43
- Mellor DH (1980) On Things and Causes in Spacetime. *The British Journal for the Philosophy of Science* 31(3):282–288
- Miller S (2008) Torture. In: Zalta EN (ed) *The Stanford Encyclopedia of Philosophy*, fall 2008 edn
- Mintzberg H (1979) *The Structuring of Organizations*. Englewood Cliffs, NJ: Prentice-Hall
- von Mises L (1966) *Human action: A treatise on economics*. H. Regnery Co.
- Morgan G (1980) Paradigms, Metaphors, and Puzzle Solving in Organization Theory. *Administrative Science Quarterly* 25(4):605–622
- Morris AD (1999) A retrospective on the civil rights movement: Political and intellectual landmarks. *Annual Review of Sociology* 25:517–539
- Mueller DC (2004) Models of Man: Neoclassical, Behavioural, and Evolutionary. *Politics, Philosophy & Economics* 3:59–76
- Muller AL (1978) Quesnay's Theory of Growth: A Comment. *Oxford Economic Papers* 30(1):150–156
- Muse CE (2009) The collapse of party politics. In: LSE Selected Pamphlets, London School of Economics and Political Science, the original pamphlet was published in 1894
- Nash J (1951) Non-cooperative games. *The Annals of Mathematics* 54(2):286–295
- von Neumann J, Morgenstern O (1956) *Theory of Games and Economic Behavior*, 3rd edn. Princeton University Press
- Nichols S, Stich SP (2003) Mindreading: An integrated account of pretence, self-awareness, and understanding of other minds. Oxford University Press
- Nietzsche F (1986) *Die Fröhliche Wissenschaft*. Kröner, 1st edn. in 1887
- O'Grady P (1999) Carnap and Two Dogmas of Empiricism. *Philosophy and Phenomenological Research* 59(4):1015–1027
- Omer Ze'ev Bekerman (2005) The Absolute Prohibition of a Relative Term: Does Everyone Know What Is in Room 101? *The American Journal of Comparative Law* 53(4):743–783
- van Osselaer SMJ, Alba JW (2000) Consumer Learning and Brand Equity. *The Journal of Consumer Research* 27(1):1–16
- Parasuraman A, Zeithaml VA, Berry LL (1985) A Conceptual Model of Service Quality and Its Implications for Future Research. *The Journal of Marketing* 49(4):41–50
- Pareto V (1900) *Sunto di alcuni capitoli di un nuovo trattato di economia pura del prof. Pareto*. *Giornale degli Economisti* 2:216–235
- Partnoy F (2003) A Revisionist View of Enron and the Sudden Death of 'May'. *Villanova Law Review* 48(4):1245
- Peirce CS (1931, 1935, 1958) *Collected Papers of Charles Sanders Peirce*. Harvard University Press
- Peirce CS (1932) *Collected Papers of Charles Sanders Peirce*. Harvard University Press
- Peirce CS (1998) *The Essential Peirce*. Indiana University Press
- Pelosi N (February 13, 2007) Speech in the United States House of Representatives: Vote Against Escalation in Iraq is a Message to President Bush – No More Blank Checks on Iraq. United States House of Representatives
- Petty W (1899) Another essay in political arithmetic. In: *The Economic Writings of Sir William Petty*, Cambridge University Press

- Anonymous (2004) Climate change activities in the United States: 2004 Update. PEW Center on Global Climate Change.
- Phillips JM (1999) Antecedents of leader utilization of staff input in decision-making teams. *Organizational Behavior and Human Decision Processes* 77(3):215–242
- Pinfield LT (1986) A field evaluation of perspectives on organizational decision making. *Administrative Science Quarterly* 31(3):365–388
- Plamper J (2001) Abolishing ambiguity: Soviet censorship practices in the 1930s. *Russian Review* 60(4):526–544
- Plato (1997) *Plato – Complete Works*. Hackett
- Plato (2003) *The Republic*. Penguin, Desmond Lee's translation
- Polanyi M (1958) *Personal Knowledge: Towards a post-critical philosophy*. University of Chicago Press
- Pollack ME (1990) Plans as Complex Mental Attitudes. In: *Intentions in Communication*, MIT Press
- Popper KR (2002) *The Open Society and Its Enemies: The Spell of Plato*, 5th edn. Routledge
- Prasad M, Perrin AJ, Bezila K, Hoffman SG, Kindleberger K, Manturuk K, Powers AS (2009) "There Must Be a Reason": Osama, Saddam, and Inferred Justification. *Sociological Inquiry* 79(2):142–162
- Priest G (1998) What is so Bad about Contradictions. *Journal of Philosophy* 95(8):410–426
- Priest G, Tanaka K (2009) Paraconsistent logic. In: Zalta EN (ed) *The Stanford Encyclopedia of Philosophy*, summer 2009 edn
- Putnam H (1975) *Mind, language, and reality*. Cambridge University Press
- Putnam H (1981) *Reason, truth, and history*. Cambridge University Press
- Putnam H (1990) *Realism with a Human Face*. Harvard University Press
- Quiggin J (1982) A theory of anticipated utility. *Journal of Economic Behavior and Organization* 3:323–343
- Raiffa H (1968) *Decision Analysis*. Addison-Wesley
- Ramsey FP (1931) Truth and probability. In: *Foundations of Mathematics and Other Logical Essays*, Kegan Paul
- Riley JC (2001) *Rising life expectancy: a global history*. Cambridge University Press
- Robbins L (1942) *An essay on the nature & significance of economic science*. Macmillan and Co. Limited, 1st edn. in 1932
- Ross L, Lepper MR, Hubbard M (1975) Perseverance in self-deception and social perception: Biased attribution processes in the debriefing paradigm. *Journal of Personality and Social Psychology* 32(5):880–892
- Ross LD (1977) The intuitive psychologist and his shortcomings: Distortions in the attribution process. In: Berkowitz L (ed) *Advances in experimental social psychology*, Academic Press
- Ross M, McFarland C, Fletcher GJO (1981) The effect of attitude on recall of past histories. *Journal of Personality and Social Psychology* 10:627–634
- Rudd K (Oct. 6, 2008) *The children of Gordon Gekko*. The Australian
- Sacks O (1998) *The Man Who Mistook His Wife for a Hat and Other Clinical Tales*. Touchstone, 1st edn. in 1970
- Savage LJ (1954) *The Foundations of Statistics*. Wiley
- Schoemaker PJH (1982) The expected utility model: Its variants, purposes, evidence and limitations. *Journal of Economic Literature* 20(2):529–563
- Schotter A (2003) Decision making with naive advice. *The American Economic Review* 93(2):196–201
- Schotter A, Sopher B (2006) Advice and behavior in intergenerational ultimatum games: An experimental approach. *Games and Economic Behavior* 58(2):365–393
- Schwitzgebel E (2008) Belief. In: Zalta EN (ed) *The Stanford Encyclopedia of Philosophy*, fall 2008 edn
- Scott N (2001) Ambiguity Versus Precision: The Changing Role of Terminology in Conference Diplomacy. In: Kurbalija J, Slavik H (eds) *Language and Diplomacy*, Mediterranean Academy of Diplomatic Studies

- Searle JR (1969) *Speech acts: An essay in the philosophy of language*. Cambridge University Press
- Searle JR (1975) A taxonomy of illocutionary acts. In: *Language, Mind, and Knowledge (Studies in the Philosophy of Science, Vol.7)*, University of Minnesota Press
- Searle JR (1983) *The Nature of Intentional States*. In: *Intentionality, an essay in the philosophy of mind*, Cambridge University Press
- Sen A (1973) Behaviour and the concept of preference. *Economica* 40(159):241–259
- Sen A (1979a) Personal utilities and public judgements: Or what's wrong with welfare economics. *The Economic Journal* 89(355):537–558
- Sen A (1979b) Utilitarianism and Welfarism. *The Journal of Philosophy* 76(9):463–489
- Sen A (1999) The possibility of social choice. *The American Economic Review* 89(3):349–378
- Shafer G (2006) *From Cournot's Principle to Market Efficiency*. Tech. rep., Rutgers University
- Shafer G, Volk V (2001) *Probability and finance: It's only a game!* Wiley
- Shafer G, Vovk V (2006) The Sources of Kolmogorov's Grundbegriffe. *Statistical Science* 21(1):70–98
- Shakespeare W (1914) *The Life of King Henry the Fifth*. In: *Complete Works of William Shakespeare*, Oxford University Press
- Sharp G (1973) *The Politics of Nonviolent Action*. Porter Sargen Publisher
- Simon HA (1955) A behavioral model of rational choice. *The Quarterly Journal of Economics* 69(1):99–118
- Slevin P, George Largner J (Nov. 15, 2001) *Bush Plan for Terrorism Trials Defined*. Washington Post p A28
- Slovic P, Tversky A (1974) Who Accepts Savage's Axiom? *Behavioral Science* 19(6):368–373
- Slovic P, Fischhoff B, Lichtenstein S (1977) Behavioral Decision Theory. *Annual Review of Psychology* 28:1–39
- Smith A (1904) *An Inquiry into the Nature and Causes of the Wealth of Nations*, 5th edn. Methuen and Co., Ltd., 1st edn. in 1776
- Sniezek JA, Buckley T (1995) Cueing and cognitive conflict in judge-advisor decision making. *Organizational Behavior and Human Decision Processes* 62(2):159–174
- Spence M (1973) Job Market Signaling. *The Quarterly Journal of Economics* 87(3):355–374
- Sperber D, Wilson D (1995) *Relevance: communication and cognition*. Wiley-Blackwell
- Starmer C (2000) Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature* 38(2):332–382
- Stenius E (1967) Mood and language game. *Synthese* 17:254–274
- Stokke A (2010) Intention-sensitive Semantics. *Synthese* 175(3):383–404
- Stoll EE (1951) Symbolism in Moby-Dick. *Journal of the History of Ideas* 12(3):440–465
- Sullivan LH (1896) The Tall Office Building Artistically Considered. *Lippincott's Magazine*
- Swedberg R (1994) Markets as Social Structures. In: Smelser NJ, Swedberg R (eds) *The Handbook of Economic Sociology*, Princeton University Press
- Swedberg R (2003) The case for an economic sociology of law. *Theory and Society* 32:1–37
- Swoyer C (2008) Properties. In: Zalta EN (ed) *The Stanford Encyclopedia of Philosophy*, fall 2008 edn
- Szabó LE (2007) Objective probability-like things with and without objective indeterminism. *Studies in the History and Philosophy of Modern Physics* 38:626–634
- Szabó LE (2009) What remains of probability? In: Dieks D, Gonzalez W, Hartmann S, Weber M, Stadler F, Uebel T (eds) *The Present Situation in the Philosophy of Science*, Springer
- Taylor A (2009) The Big Picture: Ballots, bullets and bombs in Afghanistan. URL <http://www.boston.com/bigpicture/2009/08/>
- Taylor J (1999) Review: War, Photography and Evidence. *Oxford Art Journal* 22(1):158–165
- The Norwegian Nobel Committee (2009) The Nobel Peace Prize 2009 Press Release
- Thucydides (2004) *History of the Peloponnesian War*. Dover Publications Inc., Richard Crawley's translation
- Tversky A, Kahneman D (1981) The Framing of Decisions and the Psychology of Choice. *Science* 211(4481):453–458
- Twinning JE (1980) Alienation as a social process. *The Sociological Quarterly* 21(3):417–428

- United States Congress (2008) Congressional Record. U.S. Government Printing Office, URL <http://www.gpoaccess.gov/crecord/>
- Valdés NP (1979) The Cuban revolution: Economic organization and bureaucracy. *Latin American Perspectives* 6(1):13–37
- Žižek S (1989) *The Sublime Object of Ideology*. Verso
- Wallach A (1991) Censorship in the Soviet bloc. *Art Journal* 50(3):75–83
- Wegner DM, Wenzlaff R, Kerker RM, Beattie AE (1981) Incrimination through innuendo: Can media questions become public answers? *Journal of Personality and Social Psychology* 40(5):822–832
- Wettstein HK (1984) How to Bridge the Gap between Meaning and Reference. *Synthese* 58(1): 63–84
- Williamson I, Cullingford C (1997) The Uses and Misuses of ‘Alienation’ in the Social Sciences and Education. *British Journal of Educational Studies* 45:263–275
- Wimmer H, Hogrefe GJ, Perner J (1988) Children’s understanding of informational access as source of knowledge. *Child Development* 59(2):386–396
- Wittgenstein L (1953) *Philosophical Investigations*. Blackwell
- Wittgenstein L (2001) *Tractatus Logico Philosophicus*, 2nd edn. Routledge
- Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences (USA)* 87:4576–4579
- Yaniv I, Kleinberger E (2000) Advice taking in decision making: egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes* 83(2):260–281
- Yates JF, Price PC, Lee JW, Ramirez J (1996) Good probabilistic forecasters: the “consumer’s” perspective. *International Journal of Forecasting* 12(1):41–56
- Zemach EM (1970) Four Ontologies. *The Journal of Philosophy* 8(8):231–247



# Index

## Advice

advice modeling language (AML), 224,  
245–254

advise relation, 251

a-net alphabet, 254

a-net grammar, 256

candidate solution, 253

confidence relation, 248

conflict relation, 252

consequence relation, 253

decide relation, 251

inference relation, 252

mandatory relation, 249

maximal candidate solution, 253

optional relation, 250

picosolution, 253

preference relation, 249

advisor's problem, 259–265

definition of the general, 263

resolution of, 265–272

rough formulation, 144

AML1, 236–254

alphabet of, 237, 246

a-net alphabet, 254

a-net grammar, 254

candidate solution, 245

conflict relation, 242

consequence relation, 244

grammar of, 237, 247

inference relation, 241

labeling function in, 239

maximal candidate solution, 245

picosolution, 245

semantic domain of, 238

semantic mapping in, 238

AML2, 225–236

alphabet of, 225

a-net alphabet, 254

a-net grammar, 254

candidate solution, 235

closure, 233

conflict relation, 229

consequence relation, 232

given vs. inferred information, 231

grammar of, 226

inference relation, 227

maximal candidate solution, 236

paraconsistency, 233

picosolution, 234

semantic domain of, 226

semantic mapping in, 226

a-nets, 254–257

and censorship, 30

and cheap talk, 149

and choice, 15–18

classification of, 143–145

consent to, 20

definition of, 51, 97–110

ontological commitments, 97–99

relation to a foundational ontology,  
107–110

relation to communication, 98

role of speech acts, 101–104

dictionary definition, 8

in economic coordination, 34–42

in central planning, 35–40

in markets, 40–42

essential properties of, 104–107

identity criteria for, 104–107

for the imperfectly rational, 149

intentional vagueness of, 125

meta-, 25–34, 36–40, 42

models of, 213–283

overview, 219

for the perfectly rational, 148

in political coordination, 19–34

in democracy, 21–25

in dictatorship, 20–21



- purpose of the conceptual analysis of, 49
  - qualification of, 141
  - and rational choice models, 148
  - rationale of, 159
    - analysis of, 222
    - classification of, 220
    - framework for the analysis of, 223
    - modeling of, 219
  - relevance of, 142, 143
  - role of conceptual analysis in the analysis of, 110–113
  - signs, objects, and concepts in, 78
  - taxonomy of, 196–202
    - advice-about-a-commitment concept, 200
    - advice-about-a-cue concept, 199
    - advice-about-a-goal concept, 200
    - advice-about-an-assumption concept, 200
    - advice-about-an-evaluation concept, 200
    - advice concept, 199
    - as a tool of coordination, 15, 18
    - unquestioned acceptance, 43–46
    - vagueness of, 123–132
- Advisor's problem, *see* advisor's problem, Advice
- Antireferent, 133
- Argument
  - abstract argumentation framework, 276
  - attack relation, 276
  - definition of, 275
  - justification process, 278
- Arrow, K., 170
- Bentham, J., 169
- Bernoulli, J., 164
- Capa, R., 118
- Carnap, R., 83, 84
  - distinction inside/outside a framework, 84
- Chomsky, N., 32
- Coase, R., 40, 41
- Concept, 48, 63–65
  - essential properties, 96
  - in the sign/concept/object triangle, 65–78
- Conceptual relativism, 85, 87
  - how relative is, 88
  - and primitive terms, 86
- Context
  - of reference, *see* context of, Reference relation
- Coordination mechanism, 9, 18, 43, 45
- Cournot, A., 164
- Cronin, T., 25
- Dahl, R., 26
- Damásio, A., 191
  - somatic marker hypothesis, 191
- Darnton, R., 31
- Davidson, D., 88
- Decision information
  - derived concepts
    - capability concept, 195
  - ontology, 183
    - assumption concept, 188
    - commitment concept, 194
    - cue concept, 185
    - evaluation concept, 192
    - goal concept, 189
    - ontological commitments, 174, 176, 181–182
    - role of communication, 176–182
    - scope of, 184–187
    - synthesis of, 194
- Decision-making
  - decision analysis, 150–152
    - ontology of, 150–151
  - decision problem
    - bounded rationality, 148
    - in decision analysis, 151
    - in organized anarchies, 153
    - perfect rationality, 147
  - decision rule
    - certainty, 147
    - maximize expected utility, 151
    - min-max, 147
    - in organized anarchies, 154
    - probabilistic, 147
    - satisficing, 148
  - interwoven organizational choice, 154–157
  - ontology, 146
    - interwoven organizational choice, 155–156
    - organized anarchies, 153
    - in organized anarchies, 152–154
    - rationale for a choice, 159
- Definiendum, 52–54, 58, 59, 63, 70, 110, 112
- Definiens, 52–55, 57–60, 63, 70, 110, 112
- Definition
  - caveat about the use of criteria for rigor, 60
  - circularity, 55, 59, 61, 63, 78, 91
  - conservative, 59, 60
  - criteria for rigorous, 58
  - conservativeness, 59
  - eliminability, 58–59

- guidelines for rigorous, 59
  - inferential enrichment, 59–61, 63
  - intensional, 53, 54, 58, 64, 66, 69, 70, 74, 78–80, 86, 97, 106
  - ostensive, 52–54, 64, 70, 78
  - relationship between its purpose and primitive terms, 57
- Dennett, D., 180
- Division of intellectual labor, 13
- Division of labor, 7, 10–15, 29
  
- Eco, U., 54, 64
- The Economist, magazine, 28
- Elster, J., 190
  
- Friedman, M., 25
  
- Gardenförs, P., 127
- Guarino, N., 94
  
- Hayek, F., 25, 40
- Herman, E., 32
- Hitchcock, D., 275
- Homo follis, 44
- Howard, R., 160, 161
- Hume, D., 34
  
- Indexical terms, 69, 71
- Instantiation relation, 74–76
  - vs. set membership, 74
  - vs. specialization relation, 75
  - what can and what cannot be instantiated, 75
- Intentional states
  - ascription of
    - simulation-theory, 179
    - theory-theory, 179
  - beliefs, 187
  - conception of, 177
  - desires, 187
  - evaluations, 190
  - relation to speech acts, *see* relation to intentional states, Speech acts revealed, 173
- Intergenerational ultimatum game, 16, 150
  
- Joyce, I., 35
  
- Kahneman, D., 135, 161, 169
- Keeney, R., 135, 150
- Kolmogorov, A., 162–164
- Kripke, S., 69
  
- Linnaeus, C., 44
- Lipmann, W., 45
- Locke, J., 51, 54, 78, 82
- Loewenstein, K., 20
  
- Machiavelli, N., 8, 22, 228–232, 241, 243, 244, 252, 257–259, 261, 266
- March, J., 153
- Metaphor, *see* metaphor, reference relation
- Methodological individualism, 146
- Modeling framework
  - definition of, 219
- Modeling language
  - parts of, 224
  - semantics, 224
  - syntax, 224
- Morgenstern, O., 135, 146, 175, 184
  
- Nietzsche, F., 84
  
- Object, 63–65
  - in the sign/concept/object triangle, 65–78
- Olsen, J., 153
- Ontology, 82–97
  - absolute, 84
  - depth of an, 91
  - engineered, 84–97
    - essential properties, 96
    - foundational vs. domain, 93–94
    - identity criteria, 97
    - ontological choices, *see* ontological commitments, engineered, Ontology
    - overloaded relations, 96
    - purpose of an, 91
    - scope of an, 91
    - width of an, *see* scope of an, engineered, Ontology
  - engineered vs. absolute, 84
  - engineering of, 89–97
  - ontological commitments, 82, 90, 94–97
  - evaluation of, 94–97
  - initial, 174
  - in philosophy, 83
  - in metaphysics, 83
- Orwell, G., 24

- Paraconsistency, 233
- Pareto, V., 172
- Particulars, 75, 76
- Peano, G., 53
  - Peano axioms, 53, 54
- Peirce, C., 65, 66
- Petty, W., 10–12
- Picotheory, 208–213
  - parts of, 210
    - decision information ontology, 210
    - decision problem, 211
    - decision rules, 212
    - instances of cues, 210
- Polyarchy, 26, 28–30, 32, 45
- Popper, K., 21, 22, 84
- Prima facie evidence, 86
- Primitive terms, 51, 54–58
  - and conceptual relativism, 86
  - used to postpone decisions, 56
- Probability
  - objective, 163
  - rationale for estimates of, 160
  - subjective, 163, 165
- Properties
  - monadic, 67–68
    - distinction from relational, 68
  - relational, 67–68
    - distinction from monadic, 68
- Proposition
  - concept of, 184
- Putnam, H., 85
  - on the status of truth, 87
- Quiggin, J., 169
- Raiffa, H., 135, 150
- Rationality
  - bounded, 147
  - perfect, 146
- Reference
  - mechanism of
    - causal, 69
    - contextual, 69
    - descriptive, 68
- Reference relation, 50, 68–74, 89
  - and conceptual relativism, 86
  - context of, 70–74
    - requirements for the reference relation
      - to form in context, 70–73
  - metaphor, 121–133
  - mis-reference, 122, 132–139
  - multiplicity of, *see* openness of, Reference relation
  - openness of, 117–123
    - as sign-concept relation, 76
    - use in a modeling language, 227
- Relations
  - instantiation, *see* Instantiation relation
  - reference, *see* Reference relation
  - specialization, *see* relation, specialization
- Robbins, L., 34
- Russell, B., 68
- Savage, L., 135
- Schoemaker, P., 169, 172
- Schotter, A., 16, 17, 150
- Searle, J., 174, 177, 184, 192
- Semiotics, 62
  - semiotic triangle, 64–66
- Sen, A., 170, 171
- Sharp, Gene, 28
- Sign, 48–50, 52, 53, 61–66, 68–72, 74–79, 82, 83, 85, 86, 89, 90, 98, 100, 102, 105–108, 110–114
  - definition, 62
  - medium of communicating a, 62
  - in the sign/concept/object triangle, 65–78
- Sign/concept/object triangle, 65–78
  - relations in, 76, 77
    - concept-object, *see* Instantiation relation
  - in relation to context, 70
  - use in the analysis of advice, 78
- Sign/object/concept triangle
  - and conceptual relativism, 86
- Simon, H., 135, 147, 152
- Slovic, P., 162
- Smith, A., 11–13, 34
- Specialist *vs.* generalist, 14, 15, 29, 30, 45, 47, 49
- Specialization
  - of knowledge, 7, 12–15, 45–47
  - relation, 75, 76, 91–96, 107, 108, 186, 188, 194–196, 198, 236, 246, 248, 249, 266, 268
  - of work, 7, 11–15, 18, 29, 45–47
- Specialization relation, *see* relation, specialization
- Speech acts
  - dictum/modus distinction, 99–101
  - kinds of
    - assertive, 101
    - commissive, 102
    - declarative, 103

- directive, 102
- expressive, 102
- representative declarative, 103
- relation to intentional states, 103
- role in advice, *see* role of speech acts,  
definition of, Advice
- Spence, M., 149
- Starmer, C., 169
- Szabó, L., 168
- Tversky, A., 135, 161, 162, 169
- Universals, 75, 76
- Utility
  - concept of, 168
- Vagueness
  - and conceptual spaces, 127
  - conditions for, 124
  - ontology of, 125
- Von Neumann, J., 135, 146, 175, 184
- Welty, C., 94
- Wittgenstein, L., 52, 99, 100
- The World Bank, 28