# A Theory of Case-Based Decisions

ITZHAK GILBOA

and

DAVID SCHMEIDLER

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF
THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge
United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, VIC 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org



First published 2001

Printed in the United Kingdom at the University Press
Cambridge

*Typeface* 9.75/12pt Trump Medieval *System* LaTeX 2$_\varepsilon$ [kw]

*A catalogue record for this book is available from
the British Library*

ISBN 0 521 80234 2 hardback
ISBN 0 521 00311 3 paperback

# CONTENTS

CHAPTER 1

# PROLOGUE

## 1 The scope of this book

The focus of this book is formal modeling of decision making by a single person who is aware of the uncertainty she is facing. Some of the models and results we propose may be applicable to other situations. For instance, the decision maker may be an organization or a computer program. Alternatively, the decision maker may not be aware of the uncertainty involved or of the very fact that a decision is being made. Yet, our main interest is in descriptive and normative models of conscious decisions made by humans.

There are two main paradigms for formal modeling of human reasoning, which have also been applied to decision making under uncertainty. One involves probabilistic and statistical reasoning. In particular, the Bayesian model coupled with expected utility maximization is the most prominent paradigm for formal models of decision making under uncertainty. The other employs rule-based deductive systems. Each of these paradigms provides a conceptual framework and a set of guidelines for constructing specific models for a wide range of decision problems.

These two paradigms are not the only ways in which people's reasoning may be, or has been, described. In particular, the claim that people reason by analogies dates back at least to Hume. However, reasoning by analogies has not been the subject of formal analysis to the same degree that the other paradigms have. Moreover, there is no general purpose theory we are aware of that links reasoning by analogies to decision making under uncertainty.

Our goal is to fill this gap. That is, we seek a general purpose formal model, comparable to the model of expected utility maximization, that will (i) provide a framework within which a large class of specific problems can be modeled; (ii) be based on data that are, at least in principle, observable; (iii) allow mathematical analysis of qualitative issues, such as asymptotic behavior; and (iv) be based on reasoning by analogies.

We believe that human reasoning typically involves a combination of the three basic techniques, namely, rule-based deduction, probabilistic inference, and analogies. Formal modeling tends to opt for elegance, and to focus on certain aspects of a problem at the expense of others. Indeed, our aim is to provide a model of case- or analogy-based decision making that will be simple enough to highlight main insights. We discuss the various ways in which our model may capture deductive and probabilistic reasoning, but we do not formally model the latter. It should be taken for granted that a realistic model of the human mind would have to include ingredients of all three paradigms, and perhaps several others as well. At this stage we merely attempt to lay the foundations for one paradigm whose absence from the theoretical discussion we find troubling.

The theory we present here does not purport to be more realistic than other theories of human reasoning or of choice. In particular, our goal is *not* to fine-tune expected utility theory as a descriptive theory of decision making in situations described by probabilities or states of the world. Rather, we wish to suggest a framework within which one can analyze choice in situations that do not fit existing formal models very naturally. Our theory is just as idealized as existing theories. We only claim that in many situations it is a more natural conceptualization of reality than are these other theories.

This book does not attempt to provide even sketchy surveys of the established paradigms for formal modeling of reasoning, or of the existing literature on case-based

reasoning. The interested reader is referred to standard texts for basic definitions and background.

Many of the ideas and mathematical results in this book have appeared in journal articles and working papers (Gilboa and Schmeidler 1995, 1996, 1997a,b, 1999, 2000a,b, 2001). This material has been integrated, organized, and interpreted in new ways. Additionally, several sections appear here for the first time.

In writing this book, we made an effort to address readers from different academic disciplines. Whereas several chapters are of common interest, others may address more specific audiences. The following is a brief guide to the book.

We start with two meta-theoretical sections, one devoted to definitions of philosophical terms, and the other to our own views on the way decision theory and economic theory should be conducted. These two sections may be skipped with no great loss to the main substance of the book. Yet, Section 2 may help to clarify the way we use certain terms (such as "rationality", "normative science", and the like), and Section 3 explains part of our motivation in developing the theory described in this book.

Chapter 2 of the book presents the main ideas of case-based decision theory (CBDT), as well as its formal model. It offers several decision rules, a behaviorist interpretation of CBDT, and a specification of the theory for prediction problems.

Chapter 3 provides the axiomatic foundations for the decision rules in Chapter 2. In line with the tradition in decision theory and in economics, it seeks to relate theoretical concepts to observables and to specify conditions under which the theory might be refuted.

Chapter 4, on the other hand, focuses on the epistemological underpinnings of CBDT. It compares it with the other two paradigms of human reasoning and argues that, from a conceptual viewpoint, analogical reasoning is primitive, whereas both deductive inference and probabilistic reasoning are derived from it. Whereas Chapter 3 provides the mathematical foundations of our theory, the present

chapter offers the conceptual foundations of the theory and of the language within which the mathematical model is formulated.

Chapter 5 deals with planning. It generalizes the CBDT model from a single-stage decision to a multi-stage one, and offers an axiomatic foundation for this generalization.

Chapter 6 focuses on a special case of our general model, in which the same problem is repeated over and over again. It relates to problems of discrete choice in decision theory and in marketing, and it touches upon issues of consumer theory. It also contains some results that are used later in the book.

Chapter 7 addresses questions of learning, dynamic evolution, and induction in our model. We start with an optimality result for the case of a repeated problem, which is based on a rather rudimentary form of learning. We continue to discuss more interesting forms of learning, as well as inductive inference. Unfortunately, we do not offer any profound results about the more interesting issues. Yet, we hope that the formal model we propose may facilitate discussion of these issues.

## 2 Meta-theoretical vocabulary

We devote this section to define the way we use certain terms that are borrowed from philosophy. Definitions of terms and distinctions among concepts tend to be fuzzy and subjective. The following are no exception. These are merely the definitions that we have found to be the most useful for discussing theories of decision making under uncertainty at the present state. While our definitions are geared toward a specific goal, several of them may facilitate discussion of other topics as well.

### *2.1 Theories and conceptual frameworks*

A theory of social science can be viewed as a formal mathematical structure coupled with an informal interpretation. Consider, for example, the economic theory that

consumer's demand is derived from maximizing a utility function under a budget constraint. A possible formal representation of this theory consists of two components, describing two sets, $\mathcal{C}$ and $\mathcal{P}$. The first set, $\mathcal{C}$, consists of all conceivable demand functions. A demand function maps a vector of positive prices $p \in \mathbb{R}^n_{++}$ and an income level $I \in \mathbb{R}_+$ to a vector of quantities $d(p, I) \in \mathbb{R}^n_+$, interpreted as the consumer's desired quantities of consumption under the budget constraint that total expenditure $d(p, I) \cdot p$ does not exceed income $I$. The second set, $\mathcal{P}$, is the subset of $\mathcal{C}$ that is consistent with the theory. Specifically, $\mathcal{P}$ consists of the demand functions that can be described as maximizing a utility function.[1] When the theory is descriptive, the set $\mathcal{P}$ is interpreted as all phenomena (in $\mathcal{C}$) that might actually be observed. When the theory is normative, $\mathcal{P}$ is interpreted as all phenomena (in $\mathcal{C}$) that the theory recommends. Thus, whether the theory is descriptive or normative is part of the informal interpretation.

The informal interpretation should also specify the intended applications of the theory. This is done at two levels. First, there are "nicknames" attached to mathematical objects. Thus $\mathbb{R}^n_+$ is referred to as a set of "bundles", $\mathbb{R}^n_{++}$ – as a set of positive "price vectors", whereas $I$ is supposed to represent "income" and $d$ – "demand". Second, there are more detailed descriptions that specify whether, say, the set $\mathbb{R}^n_+$ should be viewed as representing physical commodities in an atemporal model, consumption plans over time, or financial assets including contingent claims, whether $d$ denotes the demand of an individual or a household, and so forth.

Generally, the *formal structure* of a theory consists of a description of a set $\mathcal{C}$ and a description of a subset thereof, $\mathcal{P}$. The set $\mathcal{C}$ is understood to consist of conceivably observable

[1] Standard (neo-classical) consumer theory imposes additional constraints. For instance, homogeneity and continuity are often part of the definition of demand functions, and utility functions are required to be continuous, monotone, and strictly quasi-concave. We omit these details for clarity of exposition.

phenomena. It may be referred to as the *scope* of the theory. A theory thus selects a set of phenomena $\mathcal{P}$ out of the set of conceivable phenomena $\mathcal{C}$, and excludes its complement $\mathcal{C}\backslash\mathcal{P}$. What is being said about this set $\mathcal{P}$, however, is specified by the informal interpretation: it may be the *prediction* or the *recommendation* of the theory.

Observe that the formal structure of the theory does not consist of the sets $\mathcal{C}$ and $\mathcal{P}$ themselves. Rather, it consists of *formal descriptions of these sets*, $\mathcal{D}_{\mathcal{C}}$ and $\mathcal{D}_{\mathcal{P}}$, respectively. These formal descriptions are strings of characters that define the sets in standard mathematical notation. Thus, theories are not extensional. In particular, two different mathematical descriptions $\mathcal{D}_{\mathcal{P}}$ and $\mathcal{D}'_{\mathcal{P}}$ of the same set $\mathcal{P}$ will give rise to two different theories. It may be a nontrivial mathematical task to discover relationships between sets described by different theories.

It is possible that two theories that differ not only in the formal structure $(\mathcal{D}_{\mathcal{C}}, \mathcal{D}_{\mathcal{P}})$ but also in the sets $(\mathcal{C}, \mathcal{P})$ may coincide in the real world phenomena they describe. For example, consider again the paradigm of utility maximization in consumer theory. We have spelled out above one manifestation of this paradigm in the language of demand functions. But the literature also offers other theories within the same paradigm. For instance, one may define the set of conceivable phenomena to be all binary relations over $\mathbb{R}^n_+$, with a corresponding definition of the subset of these relations that conform to maximization of a real-valued function.

The informal interpretation of a theory may also be formally defined. For instance, the assignment of *nicknames* to mathematical objects can be viewed as a mapping from the formal descriptions of these objects, appearing in $\mathcal{D}_{\mathcal{C}}$ and in $\mathcal{D}_{\mathcal{P}}$, into a natural language, provided that the latter is a formal mathematical object. Similarly, one may formally define "real world phenomena" and represent the *(intended) applications* of the theory as a collection of mappings from the mathematical entities to this set. Finally, the *type of interpretation* of the theory, namely, whether it is

descriptive or normative, can easily be formalized.[2] Thus a theory may be described as a quintuple consisting of $\mathcal{D}_{\mathcal{C}}$, $\mathcal{D}_{\mathcal{P}}$, the nicknames assignment, the applications, and the type of interpretation.

We refer to the first three components of this quintuple, that is, $\mathcal{D}_{\mathcal{C}}$, $\mathcal{D}_{\mathcal{P}}$, and the nicknames assignment, as a *conceptual framework* (or *framework* for short). A conceptual framework thus describes a scope and a description of a prediction or a recommendation, and it points to a type of applications through the assignment of nicknames. But a framework does not completely specify the applications. Thus, frameworks fall short of qualifying as theories, even if the type of interpretation is given.

For instance, Savage's (1954) model of expected utility theory involves binary relations over functions defined on a measurable space. The mathematical model is consistent with real world interpretations that have nothing to do with choice under uncertainty, such as choice of streams of consumption over time, or of income profiles in a society. The nickname "space of states of the world", which is attached to the measurable space in Savage's model, defines a framework that deals with decision under uncertainty. But the conceptual framework of expected utility theory does not specify exactly what the states of the world are, or how they should be constructed. Similarly, the conceptual framework of Nash equilibrium (Nash 1951) in game theory refers to "players" and to "strategies", but it does not specify whether the players are individuals, organizations, or states, whether the theory should be applied to repeated or to one-shot situations, to situations involving few or many players, and so forth.

By contrast, the theory of expected utility maximization under risk (von-Neumann and Morgenstern 1944), as

[2] Our formal model allows other interpretations as well. For instance, it may represent a formal theory of aesthetics, where the set $\mathcal{P}$ is interpreted as defining what is beautiful. One may argue that such a theory can still be interpreted as a normative theory, prescribing how aesthetical judgment should be conducted.

well as prospect theory (Kahneman and Tversky 1979) are conceptual frameworks according to our definition. Still, they may be classified also as theories, because the scope and nicknames they employ almost completely define their applications.

*Terminological remark*: The discussion above implies that expected utility theory should be termed a framework rather than a theory. Similarly, non-cooperative games coupled with Nash equilibrium constitute a framework. Still, we follow standard usage throughout most of the book and often use "theory" where our vocabulary suggests "framework".[3] However, the term "framework" will be used only for conceptual frameworks that have several substantially distinct applications.

### *2.2 Descriptive and normative theories*

There are many possible meanings to a selection of a set $\mathcal{P}$ out of a set of conceivable phenomena $\mathcal{C}$. Among them, we find that it is crucial to focus on, and to distinguish between, two that are relevant to theories in the social sciences: descriptive and normative.

A descriptive theory attempts to describe, explain, or predict observations. Despite the different intuitive meanings, one may find it challenging to provide formal qualitative distinctions between description and explanation. Moreover, the distinction between these and prediction may not be very fundamental either. We therefore do not dwell on the distinctions among these goals.

A normative theory attempts to provide recommendations regarding what to do. It follows that normative theories are addressed to an audience of people facing decisions who are capable of understanding their recommendations. However, not every recommendation qualifies as normative science. There are recommendations that may be classified as moral, religious, or political preaching.

[3] We apply the standard usage to the title of this book as well.

These are characterized by suggesting goals to the decision makers, and, as such, are outside the realm of academic activity. There is an additional type of recommendations that we do not consider as normative theories. These are recommendations that belong to the domain of social planning or engineering. They are characterized by recommending tools for achieving pre-specified goals. For instance, the design of allocation mechanisms that yield Pareto optimal outcomes accepts the given goal of Pareto optimality and solves an engineering-like problem of obtaining it. Our use of "normative science" differs from both these types of recommendations.

A normative scientific claim may be viewed as an implicit descriptive statement about decision makers' preferences. The latter are about conceivable realities that are the subject of descriptive theories. For instance, whereas a descriptive theory of choice investigates actual preferences, a normative theory of choice analyzes the kind of preferences that the decision maker *would like* to have, that is, preferences over preferences. An axiom such as transitivity of preferences, when normatively interpreted, attempts to describe the way the decision maker would prefer to make choices. Similarly, Harsanyi (1953, 1955) and Rawls (1971) can be interpreted as normative theories for social choice in that they attempt to describe to what society one would like to belong.

According to this definition, normative theories are also descriptive. They attempt to describe a certain reality, namely, the preferences an individual has over the reality she encounters. To avoid confusion, we will reserve the term "descriptive theory" for theories that are not normative. That is, descriptive theories would deal, by definition, with "first-order" reality, whereas normative theories would deal with "second-order" reality, namely, with preferences over first-order reality. First-order reality may be external or objective, whereas second-order reality always has to do with subjective preferences that lie within the mind of an individual. Yet, first-order reality might

include actual preferences, in which case the distinction between first-order and second-order reality may become a relative matter.[4,5]

Needless to say, these distinctions are sometimes fuzzy and subjective. A scientific essay may belong to several different categories, and it may be differently interpreted by different readers, who may also disagree with the author's interpretation. For instance, the independence axiom of von-Neumann and Morgenstern's expected utility theory may be interpreted as a component of a descriptive theory. Indeed, testing it experimentally presupposes that it has a claim to describe reality. But it may also be interpreted normatively, as a recommendation for decision making under risk. To cite a famous example, Maurice Allais presented his paradox (see Allais 1953) to several researchers, including the late Leonard Savage. The latter expressed preferences in violation of expected utility theory. Allais argued that expected utility maximization is not a successful descriptive theory. Savage's reply was that his theory should be interpreted normatively, and that it could indeed help a decision maker avoid such mistakes.

Further, even when a theory is interpreted as a recommendation it may involve different types of recommendations. For instance, Shapley axiomatized his value for cooperative transferable utility games (Shapley 1953). When interpreted normatively, the axioms attempt to capture decision makers' preferences over the way in which, say, cost is allocated in different problems. A related result by

[4] There is a temptation to consider a hierarchy of preferences, and to ask which are in the realm of descriptive theories. We resist this temptation.

[5] In many cases first-order preferences would be revealed by actual choices, whereas second-order preferences would only be verbally reported. Yet, this distinction is not sharp. First, there may be first-order preferences that cannot be observed in actual choice. Second, one may imagine elaborate choice situations in which second-order preferences might be observed, as in cases where one decides on a decision-making procedure or on a commitment device.

Shapley shows that a player's value can be computed as a weighted average of her marginal contributions. This result can be interpreted in two ways. First, one may view it as an engineering recommendation: given the goal of computing a player's value, it suggests a formula for its computation. Second, one may also argue that compensating a player to the extent of her average marginal contribution is ethically appealing, and thus the formula, like the axioms, has a normative flavor.

To conclude, the distinctions between descriptive and normative theories, as well as between the latter and engineering on the one hand and preaching on the other, are not based on mathematical criteria. The same mathematical result can be part of any of these types of scientific or non-scientific activities. These distinctions are based on the author's *intent*, or on her perceived intent. It is precisely the inherently subjective nature of these distinctions that demands that one be explicit about the suggested interpretation of a theory.

It also follows that one has to have a suggested interpretation in mind when attempting to evaluate a theory. Whereas a descriptive theory is evaluated by its conformity with objective reality, a normative theory is not. On the contrary, a normative theory that suggests to people to do what they would anyway do is of questionable value. How should we evaluate a normative theory, then? Since we define normative theories to be descriptive theories dealing with second-order reality, a normative theory should also be tested according to its conformity to reality. But it is second-order reality that a normative theory should be compared to. For instance, a reasonable test of a normative theory might be whether its subjects accept its recommendations.

It is undeniable, however, that the evaluation of normative theories is inherently more problematic than that of descriptive theories. The evidence for normative theories would have to rely on introspection and self-report data much more than would the evidence for

descriptive theories. Moreover, these data may be subject to manipulation. To consider a simple example, suppose we are trying to test the normative claim that income should be evenly distributed. We are supposed to find out whether people would like to live in an egalitarian society. Simply asking people might confound their ethical preferences with their self-interest. Indeed, a person might not be sure whether she subscribes to a philosophical or a political thesis due to pure conviction or to self-serving biases. A gedanken experiment such as putting oneself behind the "veil of ignorance" (Harsanyi 1953, 1955, Rawls 1971) may assist one in finding one's own preferences, but may still be of little help in a social context. Further, the reality one tries to model, namely, a person's ethical preferences over the rules governing society, or her logico-aesthetical preferences over her own preferences are a moving target that changes constantly with actual behavior and social reality. Yet, the way we define normative theories admits a certain notion of empirical testing.

### *2.3 Axiomatizations*

In common usage, the term "axiomatization" refers to a theory. However, most axiomatizations in the literature apply to conceptual frameworks according to our definitions. In fact, the following definition of axiomatizations refers only to a formal structure $(\mathcal{D}_{\mathcal{C}}, \mathcal{D}_{\mathcal{P}})$.

An *axiomatization* of $T = (\mathcal{D}_{\mathcal{C}}, \mathcal{D}_{\mathcal{P}})$ is a mathematical model consisting of: (i) a formal structure $T' = (\mathcal{D}_{\mathcal{C}}, \mathcal{D}_{\mathcal{P}'})$, which shares the description of the set $\mathcal{C}$ with $T$, but whose set $\mathcal{P}'$ is described only in the language of phenomena that are deemed observable; (ii) mathematical results relating the set $\mathcal{P}'$ of $T'$ to the set $\mathcal{P}$ of $T$. Ideally, one would like to have conditions on observable phenomena that are necessary and sufficient for $\mathcal{P}$ to hold, namely, to have a structure $T'$ such that $\mathcal{P} = \mathcal{P}'$. In this case, $T'$ is considered to be

a "complete" axiomatization of $T$. The conditions that describe $\mathcal{P}'$ are referred to as "axioms".[6]

Observe that whether $T' = (\mathcal{D}_{\mathcal{C}}, \mathcal{D}_{\mathcal{P}'})$ is considered to be an axiomatization of $T$ depends on the question of observability of terms in $\mathcal{D}_{\mathcal{P}'}$. Consequently, the definition above will be complete only given a formal definition of "observability". We do not attempt to provide such a definition here, and we use the term "axiomatization" as if observability were well-defined.[7] Throughout the rest of this subsection we assume that the applications of the conceptual frameworks are agreed upon. We will therefore stick to standard usage and refer to axiomatizations of theories (rather than of formal structures).

Because human decisions involve inherently subjective phenomena, it is often the case that the formulation of a theory contains existential quantifiers. In this case, a complete axiomatization would also include a result regarding uniqueness. For instance, consider again the theory stating that "there exists a [so-called utility] function such that, in any decision between two choices, the consumer would opt for the one to which the [utility] function attaches a higher value". An axiomatization of this theory should provide conditions under which the consumer can indeed be viewed as maximizing a utility function in binary decisions. Further, it should address the question of uniqueness of this function: to what extent can we argue that the utility function is defined by the observable data of binary choices?

There are three reasons for which one might be interested in axiomatizations of a theory. First, the *meta-scientific* reason mentioned above: an axiomatization provides a link between the theoretical terms and the (allegedly) observable

[6] As opposed to the original meaning of the word, an "axiom" need not be indisputable or self-evident. However, evaluation of axiomatic systems typically prefers axioms that are simple in terms of mathematical formulation and transparent in terms of empirical content.

[7] Indeed, people who disagree about the definition of observability may consequently disagree whether a certain mathematical result qualifies as an axiomatization.

terms used by the theory. True to the teachings of logical positivism (Carnap 1923, see also Suppe 1974), one would like to have observable definitions of theoretical terms, in order to render the latter meaningful. Axiomatizations help us identify those theoretical differences that have observable implications, and avoid debates between theories that are observationally equivalent.

One might wish to have an axiomatization of a theory for *descriptive* reasons. Since an axiomatization translates the theory to directly observable claims, it prescribes a way to test the empirical validity of the theory. To the extent that the axioms do rule out certain conceivable observations, they also ascertain that the theory is non-vacuous, that is, falsifiable, as preached by Popper (1934). Note also that there are many situations, especially in the social sciences, where it is impractical to subject a theory to direct empirical tests. In those cases, an axiomatization can help us judge the plausibility of the theory. In this sense, axiomatizations may serve a rhetorical purpose.

Finally, one is often interested in axiomatizations for *normative* reasons. A normative theory is successful to the extent that it convinces decision makers to change the way they make their choices.[8] A set of axioms, formulated in the language of observable choice, can often convince decision makers that a certain theory has merit more than its mathematical formulation can. Thus, the normative role of axiomatizations is inherently rhetorical.

It is often the case that an axiomatization serves all three purposes. For instance, an axiomatization of utility

[8] Changing actual decision making is the ultimate goal of a normative theory. Such changes are often gradual and indirect. For instance, normative theories may first convince scientists before they sift to practitioners and to the general public. Also, a normative theory may change the way people would like to behave even if they fail to implement their stated policies, for, say, reasons of self-discipline. Finally, a normative theory may convince many people that they would like society to make certain choices, but they may not be able to bring them about for political reasons. In all of these cases the normative theories definitely have some success.

maximization in consumer theory provides a definition of the term "utility" and shows that binary choices can only define such a utility function up to an arbitrary monotone transformation. This meta-scientific exercise saves us useless debates about the choice between observationally equivalent utility functions.[9] On a descriptive level, such an axiomatization shows what utility maximization actually entails. This allows one to test the theory that consumers are indeed utility maximizers. Moreover, it renders the theory of utility maximization much more plausible, because it shows that relatively mild consistency assumptions suffice to treat consumers as if they were utility maximizers, even if they are not conscious of any utility function, of maximization, or even of the act of choice. Finally, on a normative level, an axiomatization of utility maximization may convince a person or an organization that, according to their own judgment, they should better adopt a utility function and act so as to maximize it.

Similarly, Savage's (1954) axiomatization of subjective expected utility maximization provides observable definitions of the terms "utility" and "(subjective) probability". Descriptively, it provides us with reasons to believe that there are decision makers who can be described as expected utility maximizers even if they are not aware of it, thus making expected utility maximization a more plausible theory. Finally, from a normative point of view, decision makers who shrug their shoulders when faced with the theory of expected utility maximization may find it more compelling if they are exposed to Savage's axioms.

While our use of the term "axiomatization" highlights the role of providing a link between theoretical concepts and observable data, this term is often used in other meanings. Both in mathematics and in the sciences, an "axiomatization" sometimes refers to a characterization of a certain entity by some of its properties. One is often interested in

[9] This, at least, is the standard microeconomic textbook view. For an opposing view see Beja and Gilboa (1992) and Gilboa and Lapson (1995).

such axiomatizations as decompositions of the concept that is axiomatized. Studying the "building blocks" of which a concept is made typically enhances our understanding thereof, and allows the study of related concepts. Axiomatizations in the sense used in this book will typically also have the flavor of decomposition of a construct. Thus, on top of the reasons mentioned above, one may also be interested in axiomatizations simply as a way to better understand what characterizes a certain concept and what is the driving force behind certain results. For instance, an axiomatization of utility maximization in consumer theory will often reduce utility theory to more primitive "ingredients", and will also suggest alternative theories that share only some of these ingredients, such as preferences that are transitive but not necessarily complete.

### *2.4 Behaviorist, behavioral, and cognitive theories*

We distinguish between two types of data. Behavioral data are observations of actions taken by decision makers. By contrast, cognitive data are choice-related observations that are derived from introspection, self-reports, and so forth. We are interested in actions that are open to introspection, even if they are not necessarily preceded by a deliberate decision-making process. Thus, what people say about what their choices might be, the reasons they give for actual or hypothetical choices, their recollection of and motivation for such choices are all within the category of cognitive data. In contrast to common psychological usage, we do not distinguish between cognition and emotion. Emotional motives, inasmuch as they can be retrieved by introspection and memory, are cognitive data. Other relevant data, such as certain physiological or neurological activities, will be considered neither behavioral nor cognitive.

Theories of choice can be classified according to the types of data they recognize as valid, as well as by the types of theoretical concepts they resort to. A theory is *behaviorist* if it only admits behavioral data, and if it also makes no use of

cognitive theoretical concepts. (See Watson 1913, 1930 and Skinner 1938.) We reserve the term "*behavioral*" to theories that only recognize behavioral data, but that make use of cognitive metaphors. Neo-classical economics and Savage's (1954) expected utility theory are behavioral in this sense: they only recognize revealed preferences as legitimate data, but they resort to metaphors such as "utility" and "belief". Finally, *cognitive* theories make use of cognitive data, as well as of behavior data. Typically, they also use cognitive metaphors.

Cognitive and behavioral theories often have behaviorist implications. That is, they may say something about behavior that will be consistent with some behaviorist theories but not with others. In this case, we refer to the cognitive or the behavioral theory as a *cognitive specification* of the behaviorist theories it corresponds to. One may view a cognitive specification of a behaviorist theory as a description of a mental process that *implements* the theory.

## *2.5 Rationality*

We find that purely behavioral definitions of rationality invariably miss an important ingredient of the intuitive meaning of the term. Indeed, if one adheres to the behavioral definition of rationality embodied in, say, Savage's axioms, one has a foolproof method of making rational decisions: choose *any* prior and *any* utility function, and maximize the corresponding expected utility. Adopting this method, one will never be caught violating Savage's axioms. Yet, few would accept an arbitrary choice of a prior as rational.

It follows that rationality has to do with reasoning as well as with behavior. As a first approximation we suggest the following definition. An action, or a sequence of actions is rational for a decision maker if, when the decision maker is confronted with an analysis of the decisions involved, but with no additional information, she does

not regret her choices.[10] This definition of rationality may apply not only to behavior, but also to decision processes leading to behavior. Observe that our definition presupposes a decision-making entity capable of understanding the analysis of the problems encountered.

Consider the example of transitivity of binary preferences. Many people, who exhibit cyclical preferences, regret some of their choices when exposed to this fact. For these decision makers, violating transitivity would be considered irrational. Casual observation shows that most people feel embarrassed when it is shown to them that they have fallen prey to framing effects (Tversky and Kahneman 1981). Hence we would say that, for most people, rationality dictates that they be immune to framing effects. Observe, however, that regret that follows from unfavorable resolution of uncertainty does not qualify as a test of rationality.

As another example, consider the decision maker's autonomy. Suppose that a decision maker decides on an act, $a$, and ends up choosing another act, $b$, because, say, she is emotionally incapable of forgoing $b$. If she is surprised or embarrassed by her act, it may be considered irrational. But irrationality in this example may be due to the intent to choose $a$, or to the implicit prediction that she would implement this decision. Indeed, if the decision maker knows that she is incapable of forgoing act $b$, it would be rational for her to adjust her decisions and predictions to the actual feasible set. That is, if she accepts the fact that she is technologically constrained to choose $b$, and if she so plans to do, there will be nothing irrational in this decision, and there will be no reason for her to regret not making the choice $a$, which was never really feasible. Similarly, a decision maker who has limitations in terms of simple mistakes, failing memory, limited computational capacity, and the like, may be rational as long as her decision takes these limitations into account, to the extent that they can be predicted.

[10] Alternatively, one can substitute "does not feel embarrassed by" for "does not regret".

Our definition has two properties that we find necessary for any definition of rationality and one that we find useful for our purposes. First, as mentioned above, it relies on cognitive or introspective data, as well as on behavioral data, and it cannot be applied to decision makers who cannot understand the analysis of their decisions. According to this definition it is meaningless to ask whether, say, bees, are rational. Second, it is subjective in nature. A decision maker who, despite all our preaching, insists on making frame-dependent choices, will have to be admitted into the hall of rationality. Indeed, there is too little agreement among researchers in the field to justify the hope for a unified and objective definition of rationality. Finally, our definition of rationality is closely related to the practical question of what should be done about observed violations of classical theories of choice, as we explain in the sequel. As such, we hope that this definition may go beyond capturing intuition to simplifying the discussion that follows.[11]

## *2.6 Deviations from rationality*

There is a large body of evidence that people do not always behave as classical decision theory predicts. What should we do about observed deviations from the classical notion of rational choice? Should we refine our descriptive theories or dismiss the contradicting data as exceptions that can only clutter the basic rules? Should we teach our normative theories, or modify them?

We find the definition of rationality given above useful in making these choices. If an observed mode of behavior is irrational for most people, one may suggest a normative recommendation to avoid that mode of behavior. By definition of irrationality, most people would accept this recommendation, rendering it a successful normative theory. By contrast, there is a weaker incentive to incorporate this

[11] For other views of rationality, see Arrow (1986), Etzioni (1986), and Simon (1986).

mode of behavior into descriptive theories: if these theories were known to the decision makers they describe, the decision makers would wish to change their behavior. Differently put, a descriptive theory of irrational choice is a self-refuting prophecy. If, however, an observed mode of behavior is rational for most people, they will stick to it even if theorists preach the opposite. Hence, recommending to avoid it would make a poor normative theory. But then the theorists should include this mode of behavior in their descriptive theories. This would improve the accuracy of these theories even if the theories are known to their subjects.

### *2.7 Subjective and objective terms*

Certain terms, such as "probability", are sometimes classified as subjective or objective. Some authors argue that all such terms are inherently subjective, and that the term "objective" is but a nickname for subjective terms on which there happens to be agreement. (See Anscombe and Aumann 1963.) A possible objection is raised by the following example. Five people are standing around a well that they have just found in the field. They all estimate its depth to be more than 100 feet. This is the subjective estimate of each of the five people. Yet, while they all agree on the estimate, they also all agree that there is a more objective way to measure the depth of the well.

Specifically, assume that Judy is one of the five people who have discovered the well, and that she recounts the story to her friend Jerry. Compare two scenarios. In the first scenario, Judy says, "I looked inside, and I saw that it was over 100 feet deep." In the second scenario she says, "I had dropped a stone into the well and three seconds had passed before I heard the splash." Jerry is more likely to accept Judy's estimate in the second scenario than he is in the first. We would also like to argue that Judy's estimate of the well's depth in the second scenario is more objective than in the first. This suggests a definition of objectivity