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 non-trivial 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.