CASE-BASED DECISION THEORY: FROM THE CHOICE OF ACTIONS TO REASONING ABOUT THEORIES

Jürgen Eichberger, Ani Guerdjikova

Presses de Sciences Po | « Revue économique »

2020/2 Vol. 71 | pages 283 à 306

ISSN 0035-2764
ISBN 9782724636437
DOI 10.3917/reco.712.0283

Case-Based Decision Theory: From the Choice of Actions to Reasoning about Theories

Jürgen Eichberger*
Ani Guerdjikova**

In the 1990s, David Schmeidler and Itzhak Gilboa initiated the study of decision-making under uncertainty in a completely new framework, without states but with data sets as the information on which to build choice behavior. While the first formulations of case-based decision theory (CBDT) aimed at applications in economic decision-making, this theory which takes data as a primitive concept provides an alternative foundation for deriving beliefs and driving the choice of predictions. This opened a new perspective on old questions in statistics and artificial intelligence. In this review, we summarize these developments in CBDT and highlight the immensely innovative nature of David Schmeidler's academic work.

LA THÉORIE DE LA DÉCISION AU CAS PAR CAS. DU CHOIX DES ACTIONS AU RAISONNEMENT SUR DES THÉORIES

Dans les années 1990, David Schmeidler et Itzhak Gilboa ont introduit un nouveau cadre d'analyse des décisions sous incertitude : les bases des données se substituent aux états du monde comme concept primitif du modèle et informent le choix du décideur. Au début, la théorie de la décision au cas par cas était orientée principalement vers des applications économiques, mais ses méthodes se sont avérées également pertinentes pour l'analyse des croyances et des prédictions statistiques. Cela a ouvert de nouvelles perspectives sur des questions classiques en statistique et en intelligence artificielle. Dans cet article, nous passons en revue ces développements et mettons en avant le caractère extrêmement novateur des travaux académiques de David Schmeidler.

Keywords: case-based decisions, similarity, beliefs, predictions, modes of reasoning

Mots clés: décisions au cas par cas, similarité, croyances, prédictions, modes de raisonnement

JEL Codes: C18, D80, D81.

* University of Heidelberg, Alfred Weber Institute. *Correspondence:* Bergheimer Str. 59, 69115 Heidelberg, Germany. *E-mail:* juergen.eichberger@awi.uni-heidelberg.de

** University of Grenoble-Alpes, IUF, CNRS, INRA, Grenoble INP, GAE. *Correspondence:* University of Grenoble-Alpes, 1241 rue des Résidences, 38400 Saint Martin d'Hères, France. *E-mail:* ani.guerdjikova@univ-grenoble-alpes.fr

PREFACE

Few theoretical developments in economic theory are so closely related to the fruitful cooperation of two researchers as case-based decision theory to David Schmeidler and Itzhak Gilboa. David Schmeidler could already look back at a distinguished academic career when he and his PhD student Itzhak Gilboa embarked on a novel approach to analyze decision-making under uncertainty. In the 1970s, David Schmeidler's name was associated with the study of competitive equilibrium with a continuum of traders (Schmeidler [1969]) and solution concepts in the context of cooperative game theory (Schmeidler [1972]). In the early 1980s, this work paved the way to reconsidering the theory of decision-making under uncertainty with beliefs represented by a subjective probability distribution as introduced by Savage [1954] and challenged by Ellsberg [1961] and Kahneman and Tversky [1979].

In a seminal contribution (Schmeidler [1989]), David Schmeidler provided a new paradigm for an alternative type of preference representation, Choquet expected utility (CEU) which spawned off a large number of related representations. Moreover, one of the most popular alternative representations, "maxmin expected utility" (MEU), was launched in cooperation with Itzhak Gilboa (Gilboa and Schmeidler [1989]) almost simultaneously. This earlier work on decision theory studied choice in the classical framework of a well-defined set of states of the world where the outcomes of actions would depend on the state which was actually realized. In the behaviorist tradition of revealed preferences that dominates economic theory, preferences over state-contingent outcomes are the primitive concept. Assumptions on these preferences would characterize both valuations of outcomes and beliefs as in Savage [1954].

More sensitive than most other decision theorists to the unspecified primitive concept of states and early on interdisciplinary aware of alternative approaches for choices in the face of uncertainty in artificial intelligence (e.g., Pearl [1988]), David Schmeidler and Itzhak Gilboa began to study decision-making under uncertainty in a completely new framework, without states representing the known "unknowns" but with data sets as the information on which to build choice behavior. From their previous work however, they maintained the premise of preferences as the concept on which to build representations.

While the case-based decision theory (CBDT) (Gilboa and Schmeidler [1995]) which Itzhak Gilboa and David Schmeidler initiated in the 1990s and summarized in *A Theory of Case-Based Decisions* (Gilboa and Schmeidler [2002]) still aimed at applications in economic decision-making, it became clear that this theory which takes data as we find it in innumerable data bases as a primitive concept provides an alternative foundation for deriving beliefs and driving the choice of predictions. This opened a new perspective on old questions in statistics, Bayesianism vs. frequentists, as well as on the algorithmic use of data in artificial intelligence.

There have been a couple of surveys on CBDT (Guerdjikova [2008a]) in its original interpretation as a theory about choice over actions. In the light of the more recent emphasis given to the prediction issue by Gilboa and Schmeidler [2012], we will focus on this redirection. This seems to be appropriate for a contribution to David Schmeidler's 80th birthday, highlighting his immensely innovative academic work on fundamental questions.

INTRODUCTION

In economic theory, uncertainty about the outcomes of an action is usually modeled as choice over state-contingent outcomes. In this perspective, uncertainty concerns the particular state occurring from a well-defined and perfectly known set of "states of the world." Any action leads to an outcome conditional on the realized state. It is assumed that the decision-maker can rank all actions according to a preference order. From these preferences over acts[1] one can deduce beliefs, that is subjective predictions about the occurrence of the states of the world relevant to the choice of an action. Savage [1954] provided a set of axioms for a decision-maker's preferences over actions that are equivalent to the decision-maker choosing the action according to the expected utility criterion with a subjective probability distribution representing beliefs. This subjective probability distribution can be viewed as a Bayesian prior distribution over the set of states of the world. If the situation is repeated one can update these prior distributions in the light of data generated by observing realized states. Updating a prior distribution in the light of data seems to be the only role data plays in traditional economics.

The question of how evidence from data affects decision-making, however, is much broader. Even the primitives of state-contingent decision-making, the "states" which resolve all uncertainty regarding a decision and the actions which a decision-maker considers are likely to be informed by data from past observations. Hence, it is no exaggeration to say that data sets form the core of economic theory. Statistics and decision theory suggest, however, different approaches for how to deal with data. Statistics usually presumes a stochastic process and proceeds to estimate the parameters of the process using observations from a data set. This method assumes that data is generated by a well-known type of stochastic process for which only the parameters are unknown.

Decision theory, in contrast, postulates properties of preference relations over states of the world, or states of nature. In this view, actions induce state-contingent outcomes. Rather than learning a probability distribution over states of the world by estimating a generating stochastic process, probabilities are derived from preferences and thus describe the subjective perception of uncertainty. In contrast to statistical theory, decision theory thus does not restrict beliefs to be consistent with available data. The prior is purely subjective. Consistency is required only when beliefs are updated with incoming information.[2] Only in the special case when the decision-maker is a Bayesian who learns from a prior consistent with the "true" process both approaches will be consistent and the decision-maker will behave as a statistician who eventually learns the true probability distribution.

Case-based decision theory departs from these approaches since it takes data as the primitive of the theory. Real-life decision-makers are neither

1. Savage [1954] called a state-contingent outcome "act" rather than "action." We will use both expressions interchangeably.

2. The consistency requirements can vary depending on the specific theory. The axioms of expected utility theory proposed by Savage [1954] imply dynamic consistency, consequentialism and Bayesian updating. In contrast, non-additive models use a more restricted set of conditions (Epstein and Breton [1993] and Ghirardato [2002]). Epstein and Schneider [2003], Pires [2002] and Hanany and Klibanoff [2009] provide three distinct approaches to establishing consistency requirements and axiomatizing updating rules for different classes of non-additive models.

statisticians nor are they perfectly rational and consistent in their preferences. In particular, they are not a priori endowed with a state space, and a set of actions that map states into outcomes. Moreover, real-life data are rarely organized and structured in a way that would allow for straightforward statistical analysis. Usually, the data collected differ in their accuracy, informativeness, availability and relevance to the decision at hand. Some observations are rare (possibly unique and ex ante unpredictable, e.g., "black swans") and it is not clear how to combine such rare observations with more frequent common place observations.

CBDT proposes a method for analyzing decision-making based on data directly, in particular, for situations in which statistical methods are not applicable. In the case-based decision framework, an agent makes decisions using the relevance (similarity) of past observations from the data set. Given the evidence in a data set for a problem at hand, possible past outcomes of actions are weighted according to the similarity (relevance) of the observations in which they occurred. The action with the best similarity-weighted performance is chosen. CBDT provides both practical guidance, as well as an axiomatic foundation which is important for empirically testing the theory and for estimating the subjective similarity function. For unstructured data, the specification of similarity, however, may be subjective and unrelated to the data.

More recently, CBDT has been applied to predictions based on past observations. In this context, the question of choosing the "correct" similarity function can be meaningfully addressed and one can study learning of the "correct" similarity function. From this perspective, a Bayesian can be viewed as a case-based decision-maker who learns the correct similarity function and who holds beliefs converging to the true probabilities of events, provided the underlying process is compatible with the notion of similarity. More generally, one can study the conditions under which knowledge of the correct similarity function will be useful for the decision-maker.

Finally, the language of case-based decision theory allows one also to talk about choices among theories. This meta-view can distinguish between decision-makers relying on Bayesian, or on case-based, or on rule-based reasoning. For example, one can show that, in the long run, Bayesian predictions carry more weight in structured environments with low degrees of uncertainty, whereas case-based reasoning tends to be more appropriate in complex environments.

In this survey we will proceed as follows. After introducing some leading examples, we will present the basic framework of CBDT in the second section. The third section will review some of the applications of CBDT to economic problems. In the fourth section we will focus on the contributions of CBDT to the prediction problem. Lastly, in the fifth section, we will discuss CBDT as a mode of reasoning over theories.

LEADING EXAMPLES

Before entering the more formal description of the framework, we would like to indicate the range of applications by discussing some examples illustrating the scope of decision problems case-based decision theory can address.

Example 1: *Job candidates*

Consider a CEO who seeks to hire an administrative assistant. The available acts are the various candidates for the job. The CEO does not know how well each of the candidates would perform if actually hired. A candidate may turn out to be unreliable, dishonest or incompetent. Some candidates may be very efficient at administrative tasks, but unable to deal with customers. Others might be perfect on the job, but unwilling to travel.

In this example, neither the possible outcomes, nor the states of the world are naturally implied by the description of the problem. Any attempt to specify these would require imagining every possible situation in which different characteristics of the candidate might be relevant and assigning to each such situation for each candidate an outcome.

A more realistic approach would be to ask each candidate for references, i.e., for records of past cases of employment when outcomes have been observed. To determine a utility index for each candidate, the outcomes observed in past cases are weighted by their relevance (similarity) for the decision at hand. In the basic model presented below, outcomes and similarity will be combined in order to determine the support a given past case (recommendation letter) provides for a candidate.

Example 2: *Medical treatment*

A physician examines a patient and registers her medical characteristics (blood pressure, temperature, age, medical history). The physician is considering a particular treatment and wishes to forecast the likelihood of its success. For information he has a data-base of patients with characteristics, possibly different from those of the current patient, who had been treated before. The data-base records also the outcome (success or failure) for each case.

In this example, the possible outcomes are well-defined. The relevant state space constructed from a large set of characteristics of a vast set of cases is, however, very large. Given that most of these states have never been observed, assigning probabilities to events in this state space is, in general, an impossible task.

Therefore, the physician may prefer to use the notion of similarity among past cases in order to predict the outcome in the current one. The predicted probability of success in the current case will be the weighted average of success of the treatment in past cases, where weights combine the physician's subjective similarity perception with the frequency of cases.

Example 3: *Choice between theories*

Studying a sequence of data, a scientist has to choose the theory that best explains these observations. He associates with each observed case and each theory a numerical value, which identifies the extent to which each observation supports the theory. Theories are then ranked according to the total support the data provides for them.

If the value describing the support provided by a given case for a theory is chosen to be the logarithm of the likelihood of the observation under the theory, then this method reduces to the maximum log-likelihood criterion.

These examples show that CBDT tries to address decision situations which are too unstructured and too complex to be addressed by the traditional theory of decision-making under uncertainty.

CASE-BASED DECISION THEORY

In this section, we will first present the case-based decision theory as introduced in a series of papers by Gilboa and Schmeidler ([1995], [1997a], [1997b], [2001]) and later in their book (Gilboa and Schmeidler [2002]). Then, we will provide the system of axioms which characterizes the representation, before introducing some extensions.

The General Framework

The case-based decision theory (CBDT) as suggested by Gilboa and Schmeidler [1995] models decision situations, in which neither states of the world, nor probabilities of outcomes can be naturally inferred from the description of the problem. Instead, the decision-maker (DM) is assumed to have a data base (a memory) consisting of past cases recording outcomes observed in past circumstances. For a given decision problem, alternatives are ranked in accordance to their similarity-weighted performance as recorded in the data.

We will describe the framework following Gilboa and Schmeidler ([2002], chap. 3).[3] The finite set of known *cases* is denoted by $\mathbb{C}$. The set of known possible *alternatives* is given by $\mathbb{Y}$. It is assumed that $\mathbb{Y}$ contains at least two alternatives. A *memory M* specifies for each case $c \in \mathbb{C}$ how often this case has been observed in the data. Hence, a memory is a mapping $M : \mathbb{C} \to \mathbb{Z}_0^+$. The order of occurrence of different cases is not recorded, reflecting the belief that the order of cases does not matter for the evaluation of acts.[4] Alternatively, the time component can be incorporated in the description of the problem. The set $\mathbb{M} = \left\{ M : \mathbb{C} \to \mathbb{Z}_0^+ \right\}$ denotes the set of all *hypothetical memories*.

Given a decision problem *p*, the decision-maker has to rank the *alternatives* in $\mathbb{Y}$ according to a *preference* order, which depends on the memory *M*, $\succsim_{p,M}$. Since the decision problem *p* is exogenously given and does not change, we will suppress the index *p* in the notation.

The Representation

For a given memory *M*, alternative *y* is preferred to y', $y \succsim_M y'$, if and only if

$$\sum_{c \in \mathbb{C}} M(c) v(y,c) \geq \sum_{c \in \mathbb{C}} M(c) v(y',c), \tag{1}$$

where for each case *c*, $v(y,c)$ is the *degree of support* which a single observation of case *c* provides for the choice of *y*. Intuitively, $v(y,c)$ summarizes the decision-maker's subjective judgment about the desirability of the alternative *y* based on a single observation of case *c*.

3. This framework is very similar to Gilboa and Schmeidler [2003], with the minor difference that in the former, the set of cases is finite and the data allows for repetition of cases, whereas in the latter, the set of cases is infinite, repetitions are not allowed, but for each case there is an infinite number of "equivalent" cases.

4. This invariance property appears as an explicit axiom in Billot et al. [2005].

In more specific formulations below, the degree of support can be decomposed into the perceived relevance of case c for the choice of y and the desirability of the outcome obtained in case c. An evaluation of the alternative y is obtained by aggregating these coefficients which may be positive or negative across cases, using the number of occurrences $M(c)$ of each case c as weights. This representation is unique up to an affine positive transformation, i.e., for any $y,c \in \mathbb{Y} \times \mathbb{C}$ if $v(y,c)$ represents the decision-maker's preferences, then so does $\tilde{v}(y,c) = \lambda v(y,c) + k_c$ for any $\lambda > 0$ and any $(k_c)_{c \in \mathbb{C}} \in \mathbb{R}^{\mathbb{C}}$.

Axiomatization

Representations of preferences are difficult, if not impossible, to test in experiments. An axiomatic characterization may reveal testable necessary and sufficient conditions for observable behavior. Gilboa and Schmeidler ([2002], chap. 3) provide an axiomatization for the representation 1. They assume that preferences may depend on the information about cases in the decision-maker's memory or data set. Hence, a family of preference relations over alternatives $(\succsim_M)_{M \in \mathbb{M}}$ conditional on the information in (potentially hypothetical) memories in $\mathbb{M}$ is a primitive concept of the theory.

An important property of these preferences concerns the preferential response to obtaining new information in form of an additional data set. The *combination* of two memories, M and M' results in a memory $M'' \in \mathbb{M}$ defined as the case-wise sum of observed cases, i.e., $M''(c) = M(c) + M'(c)$ for all $c \in C$. Variants of the following axioms support most axiomatizations of case-based evaluations of alternatives.

Axiom 1 (Order). For every $M \in \mathbb{M}$, $\succsim_M$ is complete and transitive.

Axiom 2 (Combination). If $y \succsim_M y'$, and $y \succsim_{M'} y'$, then $y \succsim_{M+M'} y'$.

Axiom 3 (Archimedian). If $y \succsim_M y'$, then for every $M' \in \mathbb{M}$, there exists a $k \in \mathbb{N}$ such that $y \succsim_{kM+M'} y'$.

Without Axiom 1 a real-valued representation is impossible.

Axiom 3 states that every evidence which supports y' more than y can be outweighed by a sufficient number of repetitions of cases which support y more than y'. Axiom 3 is a continuity axiom which would be violated if observations in a memory would render an alternative inferior regardless of any evidence from observing other cases. For instance, an administrative assistant who has been dishonest once may never be employed, regardless of how many additional good recommendations she would present. Similarly, the observation of a single black swan is sufficient to refute the theory "all swans are white" in favor of the theory "swans can be of different color."

Axiom 2 is a core axiom of case-based decision theory which makes an assumption on how preferences are affected by the combination of two memories or data sets. It states that if two separate pieces of evidence support the choice of y more than that of y', then so should their combination. In Example 1, if a CEO would want to hire a candidate based on each of two independent recommendations from two previous employers, she would not change her mind given the information in the combined data set. The maximal likelihood approach to the selection of theories also satisfies Axiom 2: if data set M implies that theory y has a higher likelihood than theory y' and so does data set M', then the combined data sets will also assign a higher likelihood to y than to y'. Axiom 2 is, however, less compelling in the context of hypothesis testing where two memories might both be too short in order to reject a given null hypothesis, but the combination of these memories may contain a sufficient number of observations for the hypothesis to be rejected. As Gilboa and Schmeidler [2002] point out, this is due to the inherent asymmetry between the null hypothesis, which is assumed valid until evidence to the contrary, and its rejection. Axiom 2 is also violated if similarity perceptions depend on experience (see Gilboa and Schmeidler [2003] for examples).

Axioms 1–3 are necessary but not sufficient for the existence of a representation as in Equation 1 (see Gilboa and Schmeidler [2002]). An additional axiom, which is not necessary, but which together with Axioms 1–3 guarantees Equation 1 is:

Axiom 4 (Diversity). For any four distinct alternatives, y_1, y_2, y_3 and $y_4 \in \mathbb{Y}$, there exists an $M \in \mathbb{M}$ such that $y_1 \succ_M y_2 \succ_M y_3 \succ_M y_4$. If $|\mathbb{Y}| < 4$, then for any ordering of the elements of $\mathbb{Y}$, there is a memory M such that $\succ_M$ coincides with that ordering.

Axiom 4 rules out the case that an alternative y (weakly) dominates alternative y' for all possible memories. It precludes, e.g., lexicographic preferences of the following type: a CEO working with Japanese clients might feel that it is always better to hire an assistant who speaks fluent Japanese than an assistant who does not, regardless of their letters of recommendation. In the context of prediction, it excludes the possibility that a forecast is always preferred to another one, regardless of the data.

Axioms 1–4 are sufficient for the existence of the representation and imply its uniqueness in the sense above.[5]

Extensions and Alternative Representations

There are several variations and extensions to the case-based decision model presented so far. Some of them will be discussed in this subsection. The first two extensions are useful in the context of predictions and evaluation of theories, the last one provides additional insights in the context of consumer choice.

Excluding Identical Cases

One might argue that no two cases are exactly identical as, at the very least, they differ in the time of their occurrence. If one holds this point of view, the

5. Furthermore, Axiom 4 imposes an additional linear independence condition on the values $v(y,c)$ for any four distinct acts, $y_1 \ldots y_4$. See Gilboa and Schmeidler ([2002], Theorem 3.1, 67).

previous framework appears unsatisfactory, since it requires the decision-maker to consider (at least hypothetically) any number of repetitions of any case. In response to this argument, Gilboa and Schmeidler [2003] consider an infinite set of cases, none of which can appear more than once in a data set. A data set is defined as a finite subset of the set of cases. Even though each case is unique, the decision-maker is assumed to be able to assign cases to equivalence classes, each of them with an infinite number of elements. Exchanging a case in the memory for a case in the same equivalence class leaves the decision-maker's preferences over alternatives unchanged. In this way, the representation in Equation 1 obtains under the same set of axioms adapted to take into account the new structure of the set of cases.

Ex Ante Preferences over Alternatives

The theory presented so far implicitly assumes that with no data all alternatives are considered ex ante indifferent, i.e., only data determines preferences. This assumption creates problems when the alternatives are theories ranked according to their ability to explain the data.[6] Hence, Gilboa and Schmeidler [2010] adapt the theory to allow for ex ante preferences, which are not dependent on a data set and can be interpreted as an a priori bias with respect to certain theories. For this adjustment, Axiom 2 has to be relaxed in the following way:

Axiom 2' (Recombination). If M_1, M_2, M_3 and $M_4 \in \mathbb{M}$ are such that $M_1 + M_2 = M_3 + M_4$, then there are no $y, y' \in \mathbb{Y}$ such that $y \succsim_{M_1} y'$, $y \succsim_{M_2} y'$, $y' \succsim_{M_3} y$ and $y' \succ_{M_4} y$.

This axiom is a generalization of Axiom 2 and ensures that learning is done "case-by-case." Intuitively, if two data-bases individually support the choice of y rather than y', then choosing a subset of cases that supports y' over y must mean that the rest of the cases provide support for y that more than compensates for those in support of y'.

Together with Axioms 1, 3 and 4, this leads to the following representation: $y \succsim_M y'$, if and only if

$$\sum_{c\in\mathbb{C}} M(c)v(y,c) + w(y) \geq \sum_{c\in\mathbb{C}} M(c)v(y',c) + w(y'), \tag{2}$$

where the constants $w(y)$ represent the decision-maker's ex ante ranking over the alternatives in $\mathbb{Y}$.

Differentiating between Utility and Similarity

In many applications related to consumer choice, it is useful to decompose the degree of support v into two components: *similarity* between the action under consideration and the case observed and *utility* of the outcomes recorded in cases. For this purpose, one assumes that, for a given decision problem p, each case is represented by the alternative $y_c \in \mathbb{Y}$ and the outcome $r_c \in R$ registered in case $c : c = (y; r)$. The set of cases is thus, $\mathbb{C} = \mathbb{Y} \times R$. The set of memories or

6. See Gilboa and Schmeidler [2012] and the discussion in the fifth section below ("Case-Based Reasoning about Theories").

data sets is defined as before. The representation now takes the form: $y \succsim_M y'$, if and only if $U_M(y) \geq U_M(y')$ with

$$U_M(y) = \sum_{c \in \mathbb{C}} M(c)\left[u(r_c) - \overline{u}\right] s(y, y_c). \tag{3}$$

Here $u : R \to \mathbb{R}$ is a utility function over outcomes and $\overline{u}$ denotes the decision-maker's *aspiration level*, i.e., the utility of a neutral outcome, $\overline{r}$ with $u(\overline{r}) = \overline{u}$. If all outcomes observed in the memory are neutral, the decision-maker is indifferent. Finally, $s : Y \times Y \to \mathbb{R}$ is the similarity function defined on alternatives. The value of the function s reflects the similarity of an alternative y under consideration to the alternative y_c observed in case c. Thus, the support of case c for the choice of y, $v(y,c)$ is decomposed into a similarity between the pair of alternatives $s(y, y_c)$ and the utility net of the aspiration level obtained in case c, $u(r_c) - \overline{u}$.

The concept of an aspiration level can be traced back to Simon [1957]. It formalizes the idea of satisficing behavior, i.e., the persistent choice of an alternative, which meets aspirations, as opposed to alternatives that maximize utility. E.g., a CEO who has a long memory of cases of satisfactory performance of his current administrative assistant might prefer to keep his current assistant even after seeing excellent resumes of other candidates.

The *similarity function* quantifies the decision-maker's similarity perception between the choice of act y_c observed in the memory and the choice of act y in the problem at hand. It captures the idea expressed by Hume [1758] that "from causes which appear similar we expect similar effects." For instance, a candidate y applying for a position as an administrative assistant at a magazine may present references y_c from her previous occupation with a radio station. Although the two jobs are not identical, they might be considered similar and, hence, the case y_c could be used to evaluate the candidate for the current position y. Distinct candidates may also be considered similar.

Gilboa and Wakker [2002] axiomatize Equation 3 by adding to Axioms 1–4, a fifth axiom which ensures that the relevance of a case depends only on the problem and the act, but not on the observed outcome. This property will fail if there are cases in the memory which are assigned different similarity weights depending on the outcomes observed.

CASE-BASED CHOICE: APPLICATIONS AND EXPERIMENTAL STUDIES

In this section we will briefly review applications of case-based decision theory to economic problems and report on some experimental studies on this topic.

Applications

The first applications of case-based decision theory were related to consumer theory. In this context, representation 3, which distinguishes between similarity of cases and utilities of outcomes, is of particular relevance. Two recurrent issues concern the long-run optimality of case-based decisions and the possible optimality of change-seeking behavior. These applications demonstrate that case-based decisions are usually analyzed in a dynamic context, in which decisions inform memory, while memory informs decisions. In this dynamic framework, the question of "optimal limit behavior" arises naturally.

Gilboa and Schmeidler ([2002], chap. 6) study a repeated decision problem with deterministic outcomes for each alternative. For a constant, but low aspiration level, a consumer will persistently choose an alternative which satisfies his aspirations, but does not necessarily maximize his utility. Such behavior captures the idea of "satisficing behavior" as expressed by Simon [1957]. When the aspiration level is sufficiently high, however, such that no alternatives generates positive net utility, Gilboa and Pazgal [2001] show that the decision-maker will choose each alternative with a frequency inversely proportional to its (negative) utility net of the aspiration level.[7] Such behavior can be interpreted as change-seeking. Combined with an inertia assumption in the model of Gilboa and Pazgal [2001], it can explain brand-switching behavior.

Building also on the idea of change-seeking behavior, Aragones [1997] studies the process of emergence of ideologies, i.e., of parties who adopt the same policy regardless of the state of the world. This leads to the division of society into partisan voters, who vote for their preferred ideology, and swing voters, who switch sides with every election.

More generally, Gilboa and Schmeidler ([2002], chap. 6) show that maximizing the case-based utility function 3 sequentially allows the decision-maker to obtain a unique optimum in terms of frequencies of choice. The properties of the similarity function play an important role in this process. Positive (negative) similarity between alternatives makes the choice of the more similar action less (more) desirable than the action chosen before. If acts concern consumption goods, positive (negative) similarity can be related to the consumption goods being substitutes (complements) (see Gilboa and Schmeidler [1997b]). When similarity effects are strong, consumers may be willing to forego instantaneous utility from desirable acts which are similar to acts which were chosen in the past and had delivered bad outcomes (Guerdjikova [2007]).

For the case when the aspiration level is adapted towards the latest experienced outcomes, Gilboa and Schmeidler [2001] show that a case-based decision-maker exhibits path-dependence in his reaction to prices. In particular, a consumer who derives satisfaction from the perceived value of a good net of its price will exhibit a lower willingness to buy this good after a single price increase than after several small price increases resulting in the same final price.

As already argued, optimality in the sense of choices maximizing instantaneous utility is not a general property of case-based decision-making. Jahnke, Chwolka and Simons [2005] analyze a production choice problem where firms learn the optimal price, respectively quality, decision of a monopolist. They show the sensitivity of limit behavior with respect to the specification of the model of learning.

7. As Gilboa and Schmeidler ([2002], 133) note, a high aspiration level need not imply that the alternatives bring disutility. E.g., a music lover, who prefers to listen to Beethoven and Mahler alternatingly may very well derive a lot of pleasure from music and eventually maximize his utility.

Gilboa and Schmeidler [1996] describe a process of adaptation of the aspiration level which in the limit leads to a choice of alternatives maximizing instantaneous utility. Such a process must 1) update the aspiration level upwards infinitely often in increasingly larger intervals in order to prevent the decision-maker from being suboptimally satisfied with an inferior alternative and 2) adapt the aspiration level to the maximal observed average payoff in order to avoid permanent switching at an excessively high aspiration level. Guerdjikova [2008b] extends this result to a more general class of similarity functions. Pazgal [1997] applies the same adaptation rule in the context of strategic interaction and shows that it selects a Pareto-optimal equilibrium in coordination games.

Several papers embed case-based decisions into a social learning framework. Gilboa, Postlewaite and Schmeidler [2015] show that the standard problem of utility maximization subject to a budget constraint is NP-complete. As an alternative, they propose that a consumer might use observations of the behavior of other households as a guideline for choosing a consumption bundle. For each available observation, the consumer would identify the closest consumption bundle within his budget set. To arrive at a choice, the resulting bundles would then be weighed according to the perceived similarity to each of the households. While the resulting choice can be represented as a solution of a constrained utility maximization problem with appropriately chosen constraints, the notion of optimality differs from the classical one.

An important special case of social learning occurs in networks. Blonski [1999] and Krause [2009b] model social learning in networks using different similarity functions to capture differences in social structures. Blonski [1999] examines in detail how the structure of the network combined with the aspiration level influences the learning of the optimal alternative. He shows that for a complete network, the limit choice depends on the aspiration level as well as on the share of the population choosing the optimal alternative. In the case of a star-shaped network, the choice of the central element can influence the long-run behavior of the population. Finally, in a model with δ-neighborhoods, the adoption of the optimal alternative is increasing in the size of the neighborhood δ, except when the network becomes complete and multiplicity emerges. Krause [2009b] simulates the learning process with a random network structure. He shows that for observations which are independently distributed across individuals, social learning of the optimal alternative (optimal herding) occurs. However, excessive herding may occur in scenarios where the information from others is useless (see also Krause [2009a]).

Experimental Studies

Several experimental studies find support for case-based decisions. Grosskopf, Sarin and Watson [2015] show that memory and similarity considerations play a role in one-shot decisions of a monopolist for allocating production across several markets, especially when feedback on actual past choices is not available. Ossadnik, Wilmsmann and Niemann [2013] find that in a stylized environment (choice between bets on the color of balls drawn from an urn with unknown payoffs) case-based reasoning explains behavior in 80% of the cases compared to max-min, min-max, α-max-min or reinforcement learning. Nevertheless, in terms of payoffs, modes of reasoning other than case-based decision theory perform better. Pape and Kurtz [2013] simulate case-based choices on data from psychological human classification learning experiments. They find that case-based decisions explains the data better than leading models in psychology. They fit the parameters of the model (similarity, memory, aspiration level) that best explain the data.

Bleichrodt et al. [2017] provide a methodology for identifying the similarity function from experimental data and apply it to predicting housing prices across regions in the Netherlands. They find that the only prediction of case-based decision theory that can be rejected is the Combination Axiom. This occurs when similarity has multiple dimensions and predictions might differ depending on the dimension chosen as dominant. The axiom cannot be rejected for simpler environments. As Bleichrodt et al. ([2017], 145) note, "such a violation is similar to the violations of separability over disjoint events (the sure-thing principle, or independence) found for expected utility, and is equally unsurprising."

CASE-BASED PREDICTIONS

Similar to subjective expected utility theory which neither restricts the decision-maker's subjective probability distribution nor provides any hint regarding its shape, the first version of case-based decision theory treats similarity perceptions as subjective without regard to whether they are in any sense adequate or appropriate for the problem under consideration. Indeed, in the context of individual consumption choice there is little objectivity as to what qualifies as an "optimal" choice for a subject. The definition of rationality in Gilboa and Schmeidler ([2002], 17-19) emphasizes the subjectivity of similarity even further: if a decision-maker acts in a way that he considers rational and cannot be persuaded that an alternative course of action can improve his well-being, he should be considered rational. The example of "brand-switching" behavior (Gilboa and Schmeidler [1997a]) highlights this point: presuming that each alternative has its own intrinsic value and that a consumer should consistently choose the brand with the highest value, an outside observer may deem irrational a consumer who constantly switches brands. Yet, a consumer who has preferences for variety may prefer consuming a good for a certain number of periods and switching brands once she gets tired of it. Over time such a strategy may well maximize utility.

In contrast, applying case-based decision theory in the context of predictions provides a framework where questions about the appropriateness of similarity functions can be meaningfully addressed. If alternatives are different predictions from which a decision-maker has to choose conditional on a data set, then similarity influences the likelihood of making a good prediction.

Reinterpreting the cumulative utility in the basic case-based decision model (Gilboa and Schmeidler [2002], chap. 3) as likelihood yields a model of inductive inference (Gilboa and Schmeidler [2003]) which includes well-known statistical procedures such as maximal likelihood as well as kernel estimation or kernel classification as special cases. In a similar vein, Billot et al. [2005] provide a model, in which the case-based decision-maker uses similarity-weighted frequencies of past observations in order to predict the probability distribution over outcomes.

In this section, we will discuss the two most prominent applications of case-based decision theory to the problem of prediction: Gilboa, Lieberman and Schmeidler [2006] and Billot et al. [2005].

Case-Based Predictions as Case-Based Decisions (Gilboa, Lieberman and Schmeidler [2006])

When the decision-maker has to choose from a set of alternative predictions, as in Example 2 (Medical treatment) where the physician had to make a diagnosis and choose the appropriate treatment, a case $c=(p_c;r_c)$ consists of a vector of observable characteristics, p_c, and an outcome, the correct diagnosis or prediction for this case, r_c. The decision-maker (physician) can use the observable characteristics (of the patient) in order to predict the outcome in the relevant case p. The preference representation is composed of 1) the similarity $s(p,p_c)$ between characteristics of the case under consideration p and the cases from the data set p_c, and 2) the negative of the distance between the prediction under consideration y and the outcome obtained in the case r_c, $-(r_c-y)^2$,

$$U_{p,M}(y)=-\sum_{c\in\mathbb{C}}M(c)(r_c-y)^2 s(p,p_c).$$

Gilboa, Lieberman and Schmeidler [2006] axiomatize this rule, using Axioms 1–3 together with a fourth axiom called Averaging which states that for data sets M in which only a single set of characteristics p has been observed with different realizations of outcomes r, a prediction y is preferred to y' iff y is closer to the average outcome in M, $\frac{\sum_{c\in\mathbb{C}}M(c)r_c}{\sum_{c\in\mathbb{C}}M(c)}$.

In the special case of this representation, where the set of outcomes consists of two elements, $R=\{0;1\}$, and y denotes the decision-maker's belief regarding the probability of outcome $r=1$, these four axioms are equivalent to prediction y being preferred to prediction y' iff y is closer to the similarity-weighted average in M than is y':

$$y \succsim_M y' \text{ iff}$$

$$\left| y-\frac{\sum_{c\in\mathbb{C}}s(p_c)M(c)r_c}{\sum_{c\in\mathbb{C}}s(p_c)M(c)}\right| \leq \left| y'-\frac{\sum_{c\in\mathbb{C}}s(p_c)M(c)r_c}{\sum_{c\in\mathbb{C}}s(p_c)M(c)}\right|, \tag{4}$$

where, for simplicity, we suppress the notation for the characteristics of the current case: $s(p_c)=s(p,p_c)$.

Case-Based Probabilities over Outcomes (Billot et al. [2005])

An interesting application of case-based decision-making concerns the derivation of probability distributions over outcomes from data. The representation of preferences among predictions in Equation 4 provides a link between information in the form of data and probabilistic beliefs. This link is further developed by Billot et al. [2005].[8]

Billot et al. [2005] consider a decision-maker who wishes to predict the probability distribution over outcomes. The set of alternatives is the simplex over a finite set of outcomes *R*, i.e., $\mathbb{Y} = \Delta^{|R|-1}$. Billot et al. [2005] assume that the order in which data arrive is irrelevant. Hence, each data set can be represented by a function $M \in \mathbb{M}$ as above.

Rather than applying axioms to a preference relation over predictions, Billot et al. [2005] directly study the mapping $y : \mathbb{M} \to \Delta^{|R|-1}$, which associates with each potential memory $M \in \mathbb{M}$ a prediction $y \in \mathbb{Y}$ of the decision-maker. Instead of the combination axiom (Axiom 2), Billot et al. [2005] assume a Concatenation Axiom which requires that for any $M, M' \in \mathbb{M}$, there exists an $\alpha \in (0, 1)$ such that $y(M + M') = \alpha y(M) + (1 - \alpha) y(M')$. This axiom, together with the requirement that at least three of the vectors $y(M)$ are linearly independent, ensures that $y(M)$ can be written as

$$y(M)(r) = \frac{\sum_{c \in \mathbb{C}} s(c) \hat{y}^c(r) M(c)}{\sum_{c \in \mathbb{C}} s(c) M(c)},$$

where $s(c)$ is the perceived similarity between case *c* and the current prediction, and $\hat{y}^c(r)$ denotes the probability that the decision-maker would have assigned to outcome *r* if the memory consisted of the single case *c*. Setting $\hat{y}^c(r) = d_r$ (the Dirac measure concentrated on outcome *r*), one obtains the generalization of Equation 4 to an arbitrary finite set of outcomes as a special case.

This representation allows one to view probabilities as similarity-weighted frequencies. In this context, rationality may be understood as the ability to make the best possible predictions given the data. In as far as data are generated by a process which satisfies Hume's premise that "causes which appear similar" generate "similar effects," the decision-maker's predictions will be correct in as far as his similarity judgments are aligned with those governing the data-generating process.

This result suggests that the case-based decision theory might fully resolve the issue of obtaining subjective probabilities based solely on data and without an underlying state-space. This is indeed true, when each observation in the data is compatible with a single state. Yet, for the case when observations consist of events, Gilboa and Schmeidler [2002] demonstrate that while predictions can be represented by a measure, this measure need not be non-negative.

The Concatenation Axiom proposed in Billot et al. [2005] treats frequencies independently of the number of observations on which they are based. Thus, it does not matter for the decision-maker whether the predicted probability of an outcome is based on a data set with 10 or with 1000 observations as long as the frequency of cases is the same. Eichberger and Guerdjikova [2010] modify

8. Billot et al. [2005] work with a finite set of outcomes containing at least three elements, $|R| \geq 3$. Gilboa, Lieberman and Schmeidler [2006] provide an axiomatization for $|R| = 2$, while Gilboa, Lieberman and Schmeidler [2011] extend the analysis to the case of a continuously distributed random variable.

the Concatenation Axiom by restricting it to data sets with an equal number of observations. With this modified Concatenation Axiom one obtains a set of similarity-weighted frequencies as probability distributions over outcomes. Moreover, the predicted probabilities vary with the number of observations. This generalization of Billot et al. [2005] allows one to incorporate ambiguity into case-based predictions and to model learning processes.

To test the presence of ambiguity in information conveyed by data, Arad and Gayer [2012] design an experiment in which the precision of the data observed by subjects varies. They show a dependence between the imprecision of the data and the ambiguity aversion displayed by the subjects.

A further link between case-based decisions and non-additive probabilities is provided by Gayer [2010], who shows that the use of similarity to form probabilistic judgments leads to probability-weighting functions, similar to those used in prospect theory.

Applications of Case-Based Predictions

The case-based approach to predictions and belief formation has been used in economic applications. Based on the theoretical work (Gilboa, Lieberman and Schmeidler [2006], [2011]), Gayer, Gilboa and Lieberman [2007] use housing market data in Tel Aviv to find out whether case-based reasoning by analogy to similar cases predicts real-estate prices better than rule-based reasoning. They find this hypothesis confirmed in the rental market for apartments but not for sales. Lovallo, Clarke and Camerer [2012] also compare analogy-based decisions in two empirical studies and find that case-based predictions make better forecasts.

Eichberger and Guerdjikova [2013] model decision-making under ambiguity based on available data. Decision-makers choose according to an α-max-min representation of preferences, in which beliefs combine objective characteristics of the data (number and frequency of observations) with subjective features of the decision-maker (similarity assessment of observations and perceived ambiguity).

Eichberger and Guerdjikova [2012] study the process of technological adaptation in response to a change in climate conditions. In a model with case-based decision-makers, some with optimistic and others with pessimistic attitudes towards ambiguity, both optimists and pessimists are crucial for a successful adaptation. Learning is induced by optimists, who are willing to try out new technologies for which there is little evidence available. Thus, optimists provide the public good of information, in contrast pessimists guarantee stability since they choose a technology, once adopted, persistently in the long run.

For an economy with asset markets where investors have to allocate funds between a safe and a risky asset, Eichberger and Guerdjikova [2018] study how ambiguity and ambiguity attitudes affect asset prices when consumers form expectations based on a data set of past observations. In an overlapping generations economy they describe limiting asset prices depending on the proportion of optimistic and pessimistic investor types. One can show that, with long memory, the market does not select for ambiguity neutrality. When perceived ambiguity is sufficiently small, but positive, only pessimists survive and determine prices in the long run. In contrast, with a short one-period memory, equilibrium prices are determined by Bayesians; yet, the average price of the risky asset is lower than its fundamental value.

Learning the Similarity Function and Second-Order Induction

For situations in which the data are indeed generated by an underlying similarity function, Gilboa, Lieberman and Schmeidler [2006], [2011] and Lieberman [2010] develop a method for estimating the parameters of the similarity function from data.

Second-order induction, i.e., learning the correct similarity has also been discussed more generally in the literature. For the case of i.i.d. data containing numerous observations and relatively few explanatory variables, Argenziano and Gilboa [2019] show that the learning process converges to a unique limit. However, when observations are few and there are many explanatory variables, the process has a non-unique limit and determining the correct similarity function is computationally hard. Similarly, Aragones et al. [2005] prove that identifying analogies in a data set is an NP-hard problem.

These findings can explain the use of counterfactuals (Tillio, Gilboa and Samuelson [2013]), fact-free learning, as well as the role of precedent in situations involving strategic interaction (Argenziano and Gilboa [2018]).

CASE-BASED REASONING ABOUT THEORIES

So far, we showed that the case-based decision theory provides a model for making choices and generating predictions in decision situations for which the Savage state-space model is not well adapted. Furthermore, case-based learning can lead to optimal decisions in the limit, either by appropriately adapting the aspiration level or by learning the appropriate similarity function. More recently, case-based decision-making has been applied to the problem of inductive inference over theories.

The Need for Subjectivity

The general representation in Equation 1 allows for a reinterpretation of the similarity function as a likelihood of a case in the light of a theory. When choosing among theories y, one may take the similarity between a theory y and a case c as a likelihood relation. Setting $v(y,c) = \log p(c \mid y)$ to be the logarithm of the probability of observation c given theory y implies that the decision-maker chooses the theory with the maximal likelihood given the data (see Gilboa and Schmeidler [2003]). While this specification closes the gap between case-based and statistical reasoning, it turns out that this decision rule need not lead to optimal choices in the limit.

In this spirit, Gilboa and Samuelson [2012] consider a decision-maker who applies the maximum likelihood rule in order to sequentially reject theories which do not fit the data. The remaining theories can then be used to make a prediction. When the set of potential theories is sufficiently rich, however, the maximum likelihood rule performs no better than chance: the decision-maker always finds a large set of theories that match the data, and, thus, have maximal likelihood. Yet these theories differ in their description of the future and may lead to wrong predictions. Thus, Gilboa and Samuelson [2012] argue for a subjective ex ante ordering on the set of theories, which may serve as a tie-breaker when several theories have maximal likelihood. In Gilboa and Schmeidler [2010] such an ordering and a set of axioms are provided which leads to the representation 2. The coefficients $w(y)$ of this representation can be interpreted as the ex ante subjective evaluation of theory y.

Gilboa and Schmeidler [2010] suggest to interpret these coefficients as a measure of the simplicity of the theory in the spirit of Akaike's information criterion (Akaike [1974]), or Kolmogorov's complexity measure (minimal length of the program to generate the theory's prediction, Kolmogorov [1965]), or the minimal length of description[9] (Rissanen [1978]). Among the theories with maximal likelihood for the observed sample, the decision-maker chooses the "simplest" one according to the adopted criterion. Another possible interpretation is that of a Bayesian prior,[10] with weights equal to the logarithm of the initial probability assigned to each theory.

Gilboa and Samuelson [2012] build on this idea and study the conditions necessary for learning the best theory. They find that the purely objective data-based criterion of maximum likelihood does not ensure optimal learning in the long run, neither in the deterministic nor in the stochastic case. Two forces may inhibit learning: 1) the decision-maker may be using the correct theory together with other theories, thus, making wrong predictions on average; or 2) the decision-maker may discard the correct theory, e.g., in a stochastic setting, the maximum likelihood criterion will, eventually, almost surely reject the correct theory.

In a deterministic setting, introducing a subjective order ensures continued learning when the set of theories with maximal likelihood is not a singleton. A sufficient condition for this result requires a subjective order with finite better sets. This condition is quite intuitive, since it will restrict the decision-maker to choose from a finite set if there are multiple theories with maximal likelihood. Subsequently, the decision-maker can explore this set further. If one of the theories in this set is correct, it will continue to be of maximal likelihood and will be chosen eventually, while the incorrect ones will be rejected. In contrast, if none of the theories in this set are correct, an alternative theory will eventually gain maximal likelihood and the set will be discarded in favor of another indifference class. This process will, eventually, converge to the choice of the correct theory (see Proposition 3.2 in Gilboa, Samuelson and Schmeidler [2015], 59).

In the stochastic setting, an interesting result obtains when preferences over theories y are represented by the average[11]

$$\frac{\sum_{c\in\mathbb{C}} M(c)v(y,c)}{\sum_{c\in\mathbb{C}} M(c)} + \alpha w(y),$$

9. Gilboa and Schmeidler ([2010], 1766) discuss some of the problems that arise when measuring the complexity of a theory.

10. Note however that the axiomatization does not fix the prior in a unique way (see Gilboa and Schmeidler [2010], 1766).

11. The average is taken so as to avoid that the likelihood of a theory converges to 0 as the number of observations increases.

where $v(y,c) = \log p(c \mid y)$ as before. The parameter α is the weight assigned to the subjective preference (i.e., complexity considerations or the ex ante prior). As $\alpha \to 0$, Gilboa and Samuelson [2012] show that the limit probability for the decision-maker's prediction being correct converges to the probability under the correct theory.

In the special case of a Bayesian decision-maker, who starts with a prior probability on the set of theories and uses this rule as a subjective order, either lexicographically in the deterministic case or with a vanishing weight in the stochastic case, optimal learning obtains.

The case-based decision theory challenges Bayesian reasoning and in particular its requirement for subjective assessment of probabilities even in the absence of data or in disregard of available data. Interestingly, the result of Gilboa and Samuelson [2012] shows that a certain amount of subjectivity is necessary for successful learning.

Choosing between Different Modes of Reasoning

Gilboa and Samuelson [2012] treat the case in which theories, while making different predictions, are all of the same type: they assign a probability to a sequence of observations. Gilboa, Samuelson and Schmeidler [2013] relax this condition: rather than theories, they consider conjectures, i.e., predictions that the history at a given time will belong to a certain event. Such conjectures can be assigned weights using a credence or belief function.[12]

Conjectures can be classified into several categories. Bayesian conjectures refer to a single state and can be verified at each history. Case-based conjectures do not have this property: rather, they condition their prediction on observing certain characteristics at two separate time periods, upon which the outcomes in these two periods are predicted to be identical. Thus, case-based conjectures refer to events rather than single states. Clearly, unless the specific characteristics have indeed been observed on the relevant path, a case-based prediction cannot be verified. Finally, rule-based conjectures relate the value of the observed characteristic at a given time *t* to the observed value of the outcome at that same time. They have an "if . . . then"-structure. Similarly to case-based predictions, they can be vacuous, when they only apply to certain characteristics, but not to others. They can also encompass events.

Theories or models can now be represented as combinations of conjectures of various types, where the weight of each conjecture is defined by the credence function. As information accumulates, some conjectures are rejected and assigned a weight of 0, whereas the weight assigned to the unrefuted ones is updated. Thus, the paper presents a general framework allowing to explore the decision process of a decision-maker who employs different types of conjectures to form beliefs.

12. A belief function, through its Möbius inverse, specifies a probability function on a σ-algebra of events. In the special case, where only singletons are assigned a strictly positive probability, the belief function is an additive probability (see Dempster [1967], Shafer [1967] and Jaffray [1989]).

Bayesianism versus Case-Based Reasoning

Gilboa, Samuelson and Schmeidler [2013] then ask which type of conjectures retain positive weights in the long run. Under the condition that for the set of Bayesian conjectures, the ratio of credences assigned to histories of the same length is bounded by a term that is polynomial in time, and a similar constraint for the set of case-based conjectures, the authors show that case-based reasoning will prevail with the credence assigned to Bayesian conjectures converging to 0. The clue to this result lies in the fact that the number of Bayesian conjectures increases exponentially with time, whereas the number of case-based conjectures is polynomial in time. Thus, under the restriction imposed on the weights assigned on conjectures of a given type, on almost every history, the weight of the Bayesian conjectures consistent with this history declines exponentially, whereas that of the case-based ones drops polynomially. In the limit, the decision-maker attributes all weight to case-based predictions.

Interestingly, the same result applies even for the case of an i.i.d. process, for which the decision-maker knows the probability distribution of the outcome conditional on the observed characteristic and uses this distribution to determine the relative weight of the Bayesian conjectures. As long as the decision-maker assigns a strictly positive credence to case-based reasoning (and the relative weights of case-based conjectures are bounded polynomially as above), in the limit he will reason in a case-based fashion, assigning all the weight to case-based conjectures. This result holds even if the decision-maker's Bayesian beliefs are correct. Moreover, the decision-maker will be conscious of this transition towards case-based reasoning. Notably, case-based reasoning prevails when the decision-maker faces a "large" (exponentially increasing with time) number of Bayesian conjectures, among which he cannot meaningfully discriminate. If, in contrast, the decision-maker assigns a credence close to one to a single state and the state is indeed realized, then Bayesian reasoning will remain dominant.

This result illustrates the difference between the notion of Bayesian conjectures, which are interested in predicting the exact history and can thus be refuted based on a finite number of observations, and the standard notion of a theory (also used in the stochastic setting of Gilboa and Samuelson [2012]), which concerns the limit distribution of a process and cannot be rejected with certainty based on finite histories.[13]

13. In particular, although the theory that the probability of a coin landing heads is 1/2 might be correct and the decision-maker might know this, the Bayesian conjecture for time t has to be more specific than this and explicitly state the t-period sequence of heads and tails. But the number of such sequences consistent with a limit frequency of 1/2 increases exponentially with t and only a single one is consistent with the actually observed history. At the same time, a case-based conjecture only requires the decision-maker to state whether the outcome at t will be the same (or distinct) from that at time $t' < t$. The number of such conjectures for time t is $\frac{(t-1)(t-2)}{2}$, which is a quadratic expression in t. The assumption imposed by Gilboa, Samuelson and Schmeidler [2013] on the weights of different case-based conjectures imply that the weight of the correct case-based conjecture based on the outcome at $t-1$ converges to 0 at a rate, which is at most polynomial. This gives the desired result.

Cases versus Rules

Gayer and Gilboa [2014] use a similar approach to compare rule-based and case-based conjectures. Rules correspond to deterministic theories in the language of Gilboa and Samuelson [2012] and thus make a prediction for every period. Thus, for each history a theory is either "refuted" or "unrefuted." Case-based reasoning is modeled as in Gilboa, Samuelson and Schmeidler [2013] by assigning a strictly positive credence to all simple case-based conjectures. When the process is exogenous, and the true state is one, on which some theory described by a rule is never refuted, case-based reasoning is eventually assigned 0 credence and the decision-maker learns the rule corresponding to the correct theory.

Defining three types of states: those on which the weight on case-based predictions is higher than that of rule-based from some time on, the reverse type, and the type of states on which neither mode of reasoning dominates in the long run, Gayer and Gilboa [2014] show that all three types of states are dense. Nevertheless, in a measure-theoretic sense,[14] case-based models will accrue a weight of 1 over time. This result is based on arguments similar to those establishing the predominance of case-based reasoning in the presence of Bayesian conjectures.

In contrast, when the decision-maker is predicting an endogenous process, in which observations depend on the agent's predictions, only rule-based theories will be assigned a strictly positive mass in the limit.

CONCLUSION

The theory of case-based decision-making originated as an alternative to the approach based on state-contingent outcomes (act) proposed by Savage [1954]. Modeling all possible contingencies in an uncertain situation amounts to knowing all relevant factors which might influence the outcome of an action under uncertainty. In Savage's theory, uncertainty is allowed to affect only the likelihood of events which are *known* to be relevant. This explains the well-known difficulties of Bayesian theory when updating on data which are inconsistent with the states.

One of the surprising recent developments in case-based decision theory points to its potential to deal with unforeseen contingencies. Gilboa, Minardi and Samuelson [2017] study a model of decision-making under uncertainty where the agent evaluates possible actions both by their case-based similarity and a set of "scenarios" affecting outcomes. "Scenarios" are similar to states in determining the outcomes of an action, yet they need not be mutually exclusive nor need they completely determine outcomes. Instead, the authors appeal to observable "eventualities" which link scenarios to the data of cases.

These new developments relate state-contingent outcomes in the spirit of Savage [1954] with the case-based theory of Gilboa and Schmeidler [2002]. Indeed, this new approach may help to bridge some of the inconsistencies between objective data and subjective "scenarios" involved in "unforeseen contingencies"

14. In general, the concepts of dense and meager sets are orthogonal to measure-theoretic concepts (see Marinacci [1994] for an extensive discussion of the issue).

and "undefined updates." Moreover, these new developments may be of practical use for pattern recognition techniques in artificial intelligence and deep learning, an application of case-based reasoning which we did not review in this survey (see, e.g., Hüllermeier [2007]).

REFERENCES

Akaike, H. [1974]. "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, 19 (6): 716–723.

Arad, A. and Gayer, G. [2012]. "Imprecise Data Sets as a Source of Ambiguity: A Model and Experimental Evidence," *Management Science*, 58 (1): 188–202.

Aragones, E. [1997]. "Negativity Effect and the Emergence of Ideologies," *Journal of Theoretical Politics*, 9 (2): 189–210.

Aragones, E., Gilboa, I., Postlewaite, A. and Schmeidler, D. [2005]. "Fact-Free Learning," *The American Economic Review*, 95 (5): 1355–1368.

Argenziano, R. and Gilboa, I. [2018]. "Learning What Is Similar: Precedents and Equilibrium Selection," *HEC Paris Research Paper*, ECO/SCD-2018-1264.

Argenziano, R. and Gilboa, I. [2019]. "Second-Order Induction in Prediction Problems," *PNAS*, 116: 10323–10328.

Billot, A., Gilboa, I., Samet, D. and Schmeidler, D. [2005]. "Probabilities as Similarity-Weighted Frequencies," *Econometrica*, 73 (4): 1125–1136.

Bleichrodt, H., Filko, M., Kothiyal, A. and Wakker, P. [2017]. "Making Case-Based Decision Theory Directly Observable," *American Economic Journal: Microeconomics*, 9 (1): 123–151.

Blonski, M. [1999]. "Social Learning with Case-Based Decision," *Journal of Economic Behavior and Organization*, 38 (1): 59–77.

Dempster, A. P. [1967]. "Upper and Lower Probabilities Induced by a Multivalued Mapping," *The Annals of Mathematical Statistics*, 38 (2): 325–339.

Eichberger, J. and Guerdjikova, A. [2010]. "Case-Based Belief Formation under Ambiguity," *Mathematical Social Sciences*, 60 (3): 161–177.

Eichberger, J. and Guerdjikova, A. [2012]. "Technology Adoption and Adaptation to Climate Change: A Case-Based Approach," *Climate Change Economics*, 3 (2): 1–41.

Eichberger, J. and Guerdjikova, A. [2013]. "Ambiguity, Data and Preferences for Information: A Case-Based Approach," *Journal of Economic Theory*, 148 (4): 1433–1462.

Eichberger, J. and Guerdjikova, A. [2018]. "Do Markets Prove Pessimists Right?" *International Economic Review*, 59 (4): 2259–2295.

Ellsberg, D. [1961]. "Risk, Ambiguity, and the Savage Axioms," *The Quarterly Journal of Economics*, 75 (4): 643–669.

Epstein, L. G. and Breton, M. L. [1993]. "Dynamically Consistent Beliefs Must Be Bayesian," *Journal of Economic Theory*, 61 (1): 1–22.

Epstein, L. G. and Schneider, M. [2003]. "Recursive Multiple Priors," *Journal of Economic Theory*, 113 (1): 1–31.

Gayer, G. [2010]. "Perception of Probabilities in Situations of Risk: A Case Based Approach," *Games and Economic Behavior*, 68 (1): 130–143.

Gayer, G. and Gilboa, I. [2014]. "Analogies and Theories: The Role of Simplicity and the Emergence of Norms," *Games and Economic Behavior*, 83: 267–283.

Gayer, G., Gilboa, I. and Lieberman, O. [2007]. "Rule-Based and Case-Based Reasoning in Housing Prices," *The B. E. Journal of Theoretical Economics*, 7 (1): 1284–1284.

Ghirardato, P. [2002]. "Revisiting Savage in a Conditional World," *Economic Theory*, 20 (1): 83–92.

Gilboa, I., Lieberman, O. and Schmeidler, D. [2006]. "Empirical Similarity," *The Review of Economics and Statistics*, 88 (3): 433–444.

Gilboa, I., Lieberman, O. and Schmeidler, D. [2011]. "A Similarity-Based Approach to Prediction," *Journal of Econometrics*, 162 (1): 124–131.

Gilboa, I., Minardi, S. and Samuelson, L. [2017]. "Cases and Scenarios in Decision under Uncertainty," *HEC Paris Research Paper*, ECO/SCD-2017-1200.

Gilboa, I. and Pazgal, A. [2001]. "Cumulative Discrete Choice," *Marketing Letters*, 12: 118–130.

Gilboa, I., Postlewaite, A. and Schmeidler, D. [2015]. "Consumer Choice as Constrained Imitation," *PIER Working Paper*, 15-013.

Gilboa, I. and Samuelson, L. [2012]. "Subjectivity in Inductive Inference," *Theoretical Economics*, 7 (2): 183–215.

Gilboa, I., Samuelson, L. and Schmeidler, D. [2013]. "Dynamics of Inductive Reference in a Unified Model," *Journal of Economic Theory*, 148: 1399–1432.

Gilboa, I., Samuelson, L. and Schmeidler, D. [2015]. *Analogies and Theories: Formal Models of Reasoning*. Oxford: Oxford University Press.

Gilboa, I. and Schmeidler, D. [1989]. "Maxmin Expected Utility with a Non-Unique Prior," *Journal of Mathematical Economics*, 18 (2): 141–153.

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

Gilboa, I. and Schmeidler, D. [1996]. "Case-Based Optimization," *Games and Economic Behavior*, 15 (1): 1–26.

Gilboa, I. and Schmeidler, D. [1997a]. "Act Similarity in Case-Based Decision Theory," *Economic Theory*, 9 (1): 47–61.

Gilboa, I. and Schmeidler, D. [1997b]. "Cumulative Utility Consumer Theory," *International Economic Review*, 38 (4): 737–761.

Gilboa, I. and Schmeidler, D. [2001]. "Reaction to Price Changes and Aspiration Level Adjustments," *Review of Economic Design*, 6 (2): 215–223.

Gilboa, I. and Schmeidler, D. [2002]. *A Theory of Case-Based Decisions*. Cambridge: Cambridge University Press.

Gilboa, I. and Schmeidler, D. [2003]. "Inductive Inference: An Axiomatic Approach," *Econometrica*, 71 (1): 1–26.

Gilboa, I. and Schmeidler, D. [2010]. "Likelihood and Simplicity: An Axiomatic Approach," *Journal of Economic Theory*, 145 (5): 1757–1775.

Gilboa, I. and Schmeidler, D. [2012]. *Case-Based Predictions: An Axiomatic Approach to Prediction, Classification and Statistical Learning*. Singapore: World Scientific Publishing.

Gilboa, I. and Wakker, D. S. P. [2002]. "Utility in Case-Based Decision Theory," *Journal of Economic Theory*, 105 (2): 483–502.

Grosskopf, B., Sarin, R. and Watson, E. [2015]. "An Experiment on Case-Based Decision Making," *Theory and Decision*, 79: 639–666.

Guerdjikova, A. [2007]. "Preferences for Diversification with Similarity Considerations." In Abdellaoui, M., Luce, R., Machina, M. and Munier, B. (eds). *Uncertainty and Risk: Mental, Formal, Experimental Representations*. Berlin: Springer, p. 63–83.

Guerdjikova, A. [2008a]. "Case-Based Decision Theory." In Durlauf, S. N. and Blume, L. E. (eds). *The New Palgrave Dictionary of Economics*. Basingstoke: Palgrave Macmillan, p. 1–7.

Guerdjikova, A. [2008b]. "Case-Based Learning with Different Similarity Functions," *Games and Economic Behavior*, 63 (1): 107–132.

Hanany, E. and Klibanoff, P. [2009]. "Updating Ambiguity Averse Preferences," *The B. E. Journal of Theoretical Economics*, 9 (1), https://doi.org/10.2202/1935-1704.1547.

Hüllermeier, E. [2007]. *Case-Based Approximate Reasoning*. Berlin: Springer.

Hume, D. [1748]. *An Enquiry Concerning Human Understanding*. Oxford: Clarendon Press, 2000.

Jaffray, J.-Y. [1989]. "Linear Utility Theory for Belief Functions," *Operations Research Letters*, 8 (2): 107–112.

Jahnke, H., Chwolka, A. and Simons, D. [2005]. "Coordinating Service-Sensitive Demand and Capacity by Adaptive Decision Making: An Application of Case-Based Decision Theory," *Decision Sciences*, 36 (1): 1–32.

Kahneman, D. and Tversky, A. [1979]. "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47 (2): 263–291.

Kolmogorov, A. N. [1965]. "Three Approaches to the Quantitative Definition of Information," *Probability and Information Transmission*, 1 (1): 3–11.

Krause, A. [2009a]. "Herding without Following the Herd: The Dynamics of Case-Based Decisions with Local Interactions." In Namatame, A., Kaizouji, T. and Aruka, Y. (eds). *The Complex Networks of Economic Interactions: Lecture Notes in Economics and Mathematical Systems*. Berlin: Springer, vol. 567, p. 178–190.

Krause, A. [2009b]. "Learning and Herding Using Case-Based Decision Theory," *IEEE Transactions on Systems, Man, and Cybernetics*, 39 (Part. A): 662–669.

Lieberman, O. [2010]. "Asymptotic Theory for Empirical Similarity Models," *Econometric Theory*, 26 (4): 1032–1059.

Lovallo, D., Clarke, C. and Camerer, C. [2012]. "Robust Analogizing and the Outside View: Two Empirical Tests of Case-Based Decision Making," *Strategic Management Journal*, 33 (5): 496–512.

Marinacci, M. [1994]. "Genericity: A Measure-Theoretic Analysis," *working paper*, Northwestern University.

Ossadnik, W., Wilmsmann, D. and Niemann, B. [2013]. "Experimental Evidence on Case-Based Decision Theory," *Theory and Decision*, 75 (2): 211–232.

Pape, A. and Kurtz, K. [2013]. "Evaluating Case-Based Decision Theory: Predicting Empirical Patterns of Human Classification Learning," *Games and Economic Behavior*, 82 (C): 52–65.

Pazgal, A. [1997]. "Satisficing Leads to Cooperation in Mutual Interests Games," *International Journal of Game Theory*, 26 (4): 439–453.

Pearl, J. [1988]. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco: Morgan Kaufmann Publishers.

Pires, C. P. [2002]. "A Rule for Updating Ambiguous Beliefs," *Theory and Decision*, 53 (2): 137–152.

Rissanen, J. [1978]. "Modelling by Shortest Data Description," *Automatica*, 14 (5): 465–471.

Savage, L. J. [1954]. *The Foundations of Statistics*. New York: John Wiley & Sons.

Schmeidler, D. [1969]. "Competitive Equilibria in Markets with a Continuum of Traders and Incomplete Preferences," *Econometrica*, 37 (4): 578–585.

Schmeidler, D. [1972]. "Cores of Exact Games I," *Journal of Mathematical Analysis and Applications*, 40 (1): 214–225.

Schmeidler, D. [1989]. "Subjective Probability and Expected Utility without Additivity," *Econometrica*, 57 (3): 571–587.

Shafer, G. [1967]. *A Mathematical Theory of Evidence*. Princeton: Princeton University Press.

Simon, H. A. [1957]. *Models of Man*. New York: John Wiley & Sons.

Tillio, A. D., Gilboa, I. and Samuelson, L. [2013]. "The Predictive Role of Counterfactuals," *Theory and Decision*, 74 (2): 167–182.