

Case-Based Decision Theory

Itzhak Gilboa; David Schmeidler



The Quarterly Journal of Economics, Vol. 110, No. 3 (Aug., 1995), 605-639.

Stable URL:

<http://links.jstor.org/sici?sici=0033-5533%28199508%29110%3A3%3C605%3ACDT%3E2.0.CO%3B2-H>

The Quarterly Journal of Economics is currently published by The MIT Press.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/mitpress.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

CASE-BASED DECISION THEORY*

ITZHAK GILBOA AND DAVID SCHEIDLER

This paper suggests that decision-making under uncertainty is, at least partly, case-based. We propose a model in which cases are primitive, and provide a simple axiomatization of a decision rule that chooses a "best" act based on its past performance in similar cases. Each act is evaluated by the sum of the utility levels that resulted from using this act in past cases, each weighted by the similarity of that past case to the problem at hand. The formal model of case-based decision theory naturally gives rise to the notions of satisficing decisions and aspiration levels.

In reality, all arguments from experience are founded on the similarity which we discover among natural objects, and by which we are induced to expect effects similar to those which we have found to follow from such objects. . . . From causes which appear *similar* we expect similar effects. This is the sum of all our experimental conclusions [Hume 1748].

I. INTRODUCTION

Expected utility theory enjoys the status of an almost unrivaled paradigm for decision-making in the face of uncertainty. Relying on such sound foundations as the classical works of Ramsey [1931], de Finetti [1937], von Neumann and Morgenstern [1944], and Savage [1954], the theory has formidable power and elegance, whether interpreted as positive or normative, for situations of given probabilities ("risk") or unknown ones ("uncertainty") alike.

*We are grateful to many people for conversations and discussions that influenced this work. In particular, we benefited from insightful conversations with Eva Gilboa, who also exposed us to case-based reasoning, and Akihiko Matsui and Kimberly Katz, who also referred us to Hume. A special thank-you is due to Benjamin Polak who served as a referee. His two reports on earlier versions of the paper are longer than the final product. His suggestions reshaped the paper and made it much more accessible to a reader who is not one of the authors. We are also grateful to the faculty and students of the Institute for the Learning Sciences at Northwestern University, faculty and guests at the Santa Fe Institute, as well as to Max Bazerman, Avraham Beja, Edward Green, Ehud Kalai, Morton Kamien, Edi Karni, Simon Kasif, James Peck, Stanley Reiter, Ariel Rubinstein, Michael Sang, Karl Schlag, Andrei Shleifer, Costis Skiadas, Steven Tadelis, and Amos Tversky for comments and references. Partial financial support from NSF Grants Nos. SES-9113108 and SES-9111873, the Alfred Sloan Foundation, and the Suntory Foundation are gratefully acknowledged.

While evidence has been accumulating that the theory is too restrictive (at least from a descriptive viewpoint), its various generalizations only attest to the strength and appeal of the expected utility paradigm. With few exceptions, all suggested alternatives retain the framework of the model, relaxing some of the more "demanding" axioms while adhering to the more "basic" ones. (See Machina [1987], Harless and Camerer [1994], and Camerer and Weber [1992] for extensive surveys.)

Yet it seems that in many situations of choice under uncertainty, the very language of expected utility models is inappropriate. For instance, in many decision problems under uncertainty, states of the world are neither naturally given, nor can they be simply formulated. Furthermore, often even a comprehensive list of all possible outcomes is not readily available or easily imagined. The following examples illustrate.

Example 1. As a benchmark, we first consider Savage's famous omelet problem [Savage 1954, pp. 13–15]: Savage is making an omelet using six eggs. Five of them are already opened and poured into a bowl. He is holding the sixth and has to decide whether to pour it directly into the bowl, or to pour it into a separate, clean dish to examine its freshness. This is a decision problem under uncertainty, because Savage does not know whether the egg is fresh or not. Moreover, uncertainty matters: if the egg is fresh, he will be better off pouring it directly into the bowl, saving the need to wash another dish. On the other hand, a rotten egg would result in losing the five eggs already in the bowl; thus, if the egg is not fresh, he would prefer to pour it into the clean dish.

In this example, uncertainty may be fully described by two states of the world: "the egg is fresh" and "the egg isn't fresh." Each of these states "resolves all uncertainty" as prescribed by Savage. Not only are there relatively few relevant states of the world in this example, they are also "naturally" given in the description of the problem. In particular, they can be defined independently of the acts available to the decision-maker. Furthermore, the possible outcomes can be easily defined. Thus, this example falls neatly into "decision-making under uncertainty" in Savage's model.

Example 2. A couple has to hire a nanny for their child. The available acts are the various candidates for the job. The agents do not know how each candidate would perform if hired. For instance, each candidate may turn out to be negligent or dishonest. Coming

to think about it, they realize that other problems may also occur. Some nannies are treating children well, but cannot be trusted with keeping the house in order. Others appear to be just perfect on the job, but are not very loyal and may quit the job on short notice.

The couple is facing uncertainty regarding the candidates' performance on several measures. However, there are a few difficulties in fitting this problem into the framework of expected utility theory (EUT). First, imagining all possible outcomes is not a trivial task. Second, the "states of the world" do not naturally suggest themselves in this problem. Furthermore, if the agents should try to construct them analytically, their number and complexity would be daunting: every state of the world should specify the exact performance of each candidate on each measure.

Example 3. President Clinton has to decide on military intervention in Bosnia-Herzegovina. (A problem that he is facing while this paper is being written, revised, and re-revised.) The alternative acts are relatively clear: one may do nothing; impose economic sanctions; use limited military force (say, only air strikes), or opt for a full-blown military intervention. Of course, the main problem is to decide what are the likely short-run and long-run outcomes of each act. For instance, it is not exactly clear how strong are the military forces of the warring factions in Bosnia; it is hard to judge how many casualties each military option would involve, and what would be the public opinion response; there is some uncertainty about the reaction of Russia, especially if it goes through a military coup.

In short, the problem is definitely one of decision under uncertainty. But, again, neither all possible eventualities, nor all possible scenarios are readily available. Any list of outcomes or of states is bound to be incomplete. Furthermore, each state of the world should specify the result of each act at each point of time. Thus, an exhaustive set of the states of the world certainly does not naturally pop up.

In example 1, expected utility theory seems a reasonable description of how people think about the decision problem. By contrast, we argue that in examples such as 2 and 3, EUT does not describe a plausible cognitive process. Should the agent attempt to "think" in the language of EUT, she would have to imagine all possible outcomes and all relevant states. Often the definition of a state of the world would involve conditional statements, attaching

outcomes to acts. Not only would the number of states be huge, the states themselves would not be defined in an intuitive way.

Moreover, even if the agent managed to imagine all outcomes and states, her task would by no means be done. Next she would have to assess the utility of each outcome, and to form a prior over the state space. It is not clear how the utility and the prior are to be defined, especially since past experience appears to be of limited help in these examples. For instance, what is the probability that a particular candidate for the job in example 2 will end up being negligent? Or being both negligent and dishonest? Or, considering example 3, what are the chances that a military intervention will develop into a full-blown war, while air strikes will not? What is the probability that a scenario that no expert predicted will eventually materialize?

It seems unlikely that decision-makers can answer these questions. Expected utility theory does not describe the way people "really" think about such problems. Correspondingly, it is doubtful that EUT is the most useful tool for predicting behavior in applications of this nature. A theory that will provide a more faithful description of how people think would have a better chance of predicting what they will do. How do people think about such decision problems, then? We resort to Hume [1748], who argued that "From causes which appear *similar* we expect similar effects. This is the sum of all our experimental conclusions." That is, the main reasoning technique that people use is drawing analogies between past cases and the one at hand.¹

Applying this idea to decision-making, we suggest that people choose acts based on their performance in similar problems in the past. For instance, in example 2 a common, and indeed very reasonable, thing to do is to ask each candidate for references. Every recommendation letter provided by a candidate attests to his/her performance (as a nanny) in a different problem. In this example, the agents do not rely on their own memory; rather, they draw on the experience of other employers. Each past "case" would be judged for its similarity; for instance, serving as a nanny to a month-odd toddler is somewhat different from the same job when a

1. We were first exposed to this idea as an explicit theory in the form of case-based reasoning [Schank 1986; Riesbeck and Schank 1989], to which we owe the epithet "case-based." Needless to say, our thinking about the problem was partly inspired by case-based reasoning. At this early stage, however, there does not seem to be much in common—beyond Hume's basic idea—between our theory and case-based reasoning. It should be mentioned that similar ideas were also expressed in the economics literature by Keynes [1921], Selten [1978], and Cross [1983].

two-year-old child is concerned. Similarly, the house, the neighborhood, and other factors may affect the relevance of past cases to the problem at hand. Thus, we expect the agents to put more weight on the experience of people whose decision problem was "more similar" to theirs. Furthermore, they may rely more heavily on the experience of people they happen to know, or judge to have tastes similar to their own.

Next consider example 3. While military and political experts certainly do try to write down possible "scenarios" and to assign likelihood to them, this is by no means the only reasoning technique used. (Nor is it necessarily the most compelling a priori or the most successful a posteriori.) Very often the reasoning used is by analogies to past cases. For instance, proponents of military intervention tend to cite the Gulf War as a "successful" case. They stress the similarity of the two problems, say, as local conflicts in post-cold-war world. Opponents adduce the Vietnam War as a case in which military intervention is generally considered to have been a mistake. They also point to the similarity of the cases, for instance to the "peace-keeping mission" mentioned in both.

Specifically, we suggest the following theory, which we dub "case-based decision theory" (CBDT). Assume that a set of "problems" is given as primitive, and that there is some measure of similarity on it. The problems are to be thought of as descriptions of choice situations, as "stories" involving decision problems. Generally, an agent would remember some of the problems that she and other agents encountered in the past. When faced with a new problem, the similarity of the situation brings this memory to mind, and with it the recollection of the choice made and the outcome that resulted. We refer to the combination of these three—the problem, the act, and the result—as a case. Thus, "similar" cases are recalled, and based on them each possible decision is evaluated. The specific model we propose and axiomatize here evaluates each act by the sum, over all cases in which it was chosen, of the product of the similarity of the problem to the one at hand and the resulting utility. (Utility will be assumed scaled such that zero is a default value.)

Formally, a *case* is a triple (q,a,r) , where q is a problem, a is an act, and r a result.² Let M , the memory, be a set of such cases. A

2. We implicitly assume that the description of a problem includes the specification of available acts. In particular, we do not address here the problem of identifying which acts are, indeed, available in a given problem, or identifying a decision problem in the first place.

decision-making agent is characterized by a *utility* function u , which assigns a numerical value to results, and a *similarity* function s , which assigns nonnegative values to pairs of problems. When faced with a new problem p , our agent would choose an act a which maximizes

$$(*) \quad U(a) = U_{p,M}(a) = \sum_{(q,a,r) \in M} s(p,q)u(r),$$

where the summation over the empty set is taken to yield zero.

In CBDT, as in EUT, acts are ranked by weighted sums of utilities. Indeed, this formula so resembles that of expected utility theory that one may suspect CBDT to be no more than EUT in a different guise. However, despite appearances, the two theories have little in common. First, note some mathematical differences between the formulae. In CBDT there is no reason for the coefficients $s(p, \cdot)$ to add up to 1 or to any other constant. More importantly, while in EUT every act is evaluated at *every* state, in CBDT each act is evaluated over a different set of cases. To be precise, if $a \neq b$, the set of elements of M summed over in $U(a)$ is disjoint from that corresponding to $U(b)$. In particular, this set may well be empty for some a 's.

On a more conceptual level, in expected utility theory the set of states is assumed to be an exhaustive list of all possible scenarios. Each state "resolves all uncertainty," and, in particular, attaches a result to each available act. By contrast, in case-based decision theory the memory contains only those cases that actually happened. Each case provides information only about the act that was chosen in it, and the evaluation of this act is based on the actual outcome that resulted in this case. Hence, to apply EUT, one needs to engage in hypothetical reasoning, namely to consider all possible states and the outcome that would result from each act in each state. To apply CBDT, no hypothetical reasoning is required.

As opposed to expected utility theory, CBDT does not distinguish between "certain" and "uncertain" acts. In hindsight, an agent may observe that a particular act always resulted in the same outcome (i.e., that it seems to "involve no uncertainty"), or that it is uncertain in the sense that it resulted in different outcomes in similar problems. But the agent is not assumed to "know" a priori which acts involve uncertainty and which do not. Indeed, she is not assumed to know anything about the outside world, apart from past cases.

CBDT and EUT also differ in the way they treat new informa-

tion and evolve over time. In EUT new information is modeled as an event, i.e., a subset of states, which has obtained. The model is restricted to this subset, and the probability is updated according to Bayes' rule. By contrast, in CBDT new information is modeled primarily by adding cases to memory. In the basic model, the similarity function calls for no update in the face of new information. Thus, EUT implicitly assumes that the agent was born with knowledge of and beliefs over all possible scenarios, and her learning consists of ruling out scenarios which are no longer possible. On the other hand, according to CBDT, the agent was born completely ignorant, and she learns by expanding her memory. (In the sequel we will also briefly discuss learning that is reflected in changes of the similarity judgments.) Roughly, an EUT agent learns by observing what cannot happen, whereas a CBDT agent learns by observing what can.

The framework of CBDT provides a natural way to formalize both the idea of frequentist belief formation (insofar as it is reflected in behavior) and the idea of satisficing. Although beliefs and probabilities do not explicitly exist in this model, in some cases they may be implicitly inferred from the number of summands in (*). That is, if the decision-maker happens to choose the same act in many similar cases, the evaluation function (*) may be interpreted as gathering statistical data, or as forming a frequentist prior. However, CBDT does not presuppose any *a priori* beliefs. Actual cases generate statistics, but no beliefs are assumed in the absence of data.

If an agent faces similar problems repeatedly, it is natural to evaluate an act by its average past performance, rather than by a mere summation as in (*). Both decision criteria can be thought of as performing "implicit" induction: they are ways to learn from past cases which decision should be made in a new problem. A case-based decision-maker does not explicitly formulate "rules." She could never arrive at any "knowledge" regarding the future. (Indeed, this is also in line with Hume's teachings.) But she may come to behave as if she realized, or at least believed in certain regularities.

Case-based decisions may result in conservative or uncertainty-averse behavior. For example, if each act $a \in A$ only ever results in a particular outcome r_a , then the agent will only try new acts until she finds one that yields $u(r_a) > 0$. Thereafter, she will choose this act over and over again. She will be satisfied with the "reasonable"

act a (so defined by $U(a) > 0$), and will not attempt to maximize her utility function u . Thus, CBDT has some common features with the notion of "satisficing" decisions of Simon [1957] and March and Simon [1958], and may be viewed as formalizing this idea. Specifically, the number zero on the utility scale may be interpreted as the agent's "aspiration level": so long as it is not reached, she keeps experimenting; once this level is obtained, she is satisfied.

Further discussion may prove more useful after a formal presentation of our model, axioms, and results. We devote Section II to this purpose. In Section III we discuss the model and its axiomatization. Further discussion, focusing on the comparison of CBDT to EUT, is relegated to Section IV. Section V presents some economic applications. In Section VI we suggest some variations on the basic theme, and discuss avenues for further research.

II. THE MODEL

Let P and A be finite and nonempty sets, of *problems* and of *acts*, respectively. To simplify notation, we will assume that all the acts A are available at all problems $p \in P$. It is straightforward to extend the model to deal with the more general case in which for each $p \in P$ there is a subset $A_p \subseteq A$ of available acts. Let R be a set of *outcomes* or *results*. For convenience, we include in R an outcome r_0 to be interpreted as "this act was not chosen." The set of *cases* is $C \equiv P \times A \times R$.

Given a subset of cases $M \subseteq C$, denote its projection on P by H . That is,

$$H = H(M) = \{q \in P \mid \exists a \in A, r \in R, \text{ such that } (q, a, r) \in M\}.$$

H will be referred to as the *history* of problems.

A *memory* is a subset $M \subseteq C$ such that

- (i) for every $q \in H(M)$ and $a \in A$, there exists a unique $r = r_M(q, a)$ such that $(q, a, r) \in M$;
- (ii) for every $q \in H(M)$ there is a unique $a \in A$ for which $r_M(q, a) \neq r_0$.

A memory M may be viewed as a function, assigning results to pairs of the form (problem, act). For every memory M , and every $q \in H = H(M)$, there is one act that was actually chosen at q —with an outcome $r \neq r_0$ defined by the past case—and the other acts will be assigned r_0 .

The definition of memory makes two implicit simplifying assumptions, which entail no loss of generality: first, we assume that no problem $p \in P$ may be encountered more than once. However, the fact that two formally distinct problems may be "practically identical" (as far as the agent is concerned) can be reflected in the similarity function. Second, we define memory to be a *set*, implying that the *order* in which cases appear in memory is immaterial. Yet, if the description of a problem is informative enough, for instance, if it includes a time parameter, a set is as informative as a sequence.

To simplify exposition, we will henceforth assume (explicitly) that $R = \mathcal{R}$ (the reals) and (implicitly) that it is already measured in "utils." That is, our axioms should be interpreted as if R were scaled so that the "utility" function be the identity. Furthermore, we will assume that $r_0 = 0$. (See Section III for a discussion of these assumptions.) We do not distinguish between the actual outcome 0 and r_0 . In particular, it is possible that for some $q \in H(M)$, $r_M(q, a) = 0 = r_0$ for all $a \in A$.

Though by no means necessary, it may be helpful to visualize a memory, which is a function from $A \times H$ to \mathcal{R} , as a matrix. That is, choosing arbitrary orderings of A and of $H = H(M)$, a memory M can also be thought of as a $(k \times n)$ -real-valued matrix, in which the $k \equiv |A|$ rows correspond to acts, and the $n \equiv |H|$ columns—to problems in H . In such a matrix every column contains at most one nonzero entry. Conversely, every $(k \times n)$ -matrix which satisfies this condition corresponds to some memory M' with $H(M') = H$. Thus, every such matrix may be viewed as a *conceivable* memory, which may differ from the actual one in terms of the acts chosen at the various problems, as well as the results they yielded.

We assume that, when the agent has memory M and is confronted with problem p , she chooses an act in accordance with a preference relation $\geq_{p, M} \subseteq A \times A$. We further assume that the evaluation of an act is based only on the outcomes which resulted from the act. This assumption has two implications. First, for a given memory, each act may be identified with its "act profile," that is, with a vector in \mathcal{R}^H , specifying the results it yielded in past problems. Thus, a memory matrix M induces a preference order over k vectors in \mathcal{R}^H , namely, its rows.

Second, we require that the preference between two real-valued vectors not depend on the memory which contains them. Formally, for $x, y \in \mathcal{R}^H$, assume that M and M' are such that $H(M) = H(M') = H$, and that each of x and y corresponds to a row

in the matrix M and to a row in the matrix M' . Then we require that $x \geq_{p,M} y$ iff $x \geq_{p,M'} y$.

Under these assumptions we can simply postulate a preference order $\geq_{p,H}$ on \mathcal{A}^H , which depends only on p and the observed problems H ($p \notin H$). One interpretation of this preference order is that the agent can not only rank acts given their actual profile, but also provide preferences among hypothetical act profiles. (See a discussion of this point in the following section.)

However, we will not assume that $\geq_{p,H}$ is a complete order on \mathcal{A}^H . Consider two distinct act profiles $x, y \in \mathcal{A}^H$, assigning $x(q) \neq 0$ and $y(q) \neq 0$, respectively, to some $q \in H$. Naturally, these cannot be compared even hypothetically: for any memory M , at most one act may be chosen in problem q , and therefore at most one act may have a value different from 0 in its act profile for any given q . In other words, there is no memory matrix in which both x and y appear as rows. We therefore restrict the partial order $\geq_{p,H}$ to compare act profiles which are *compatible* in the sense that they could appear in the same memory matrix. Formally, given $x, y \in X$, let $x*y \in \mathcal{A}^H$ be defined as a coordinatewise product; i.e., $(x*y)(q) = x(q)y(q)$ for $q \in H$. Using this notation, two act profiles x, y are compatible if $x*y = 0$ or $x = y$.

Our first axiom states that compatibility is necessary and sufficient for comparability. Since compatibility is not a transitive relation, this axiom implies that neither is $\geq_{p,H}$.

A1. COMPARABILITY OF COMPATIBLE PROFILES. For every $p \in P$ and every history $H = H(M)$, for every $x, y \in \mathcal{A}^H$, x and y are compatible iff $x \geq_{p,H} y$ or $y \geq_{p,H} x$.

The following three axioms will guarantee the additively separable representation of $\geq_{p,H}$ on \mathcal{A}^H .

A2. MONOTONICITY. For every p, H , $x \geq y$ and $x*y = 0$ implies that $x \geq_{p,H} y$.

A3. CONTINUITY. For every p, H , and $x \in \mathcal{A}^H$, the sets $\{y \in \mathcal{A}^H \mid y \geq_{p,H} x\}$ and $\{y \in \mathcal{A}^H \mid x \geq_{p,H} y\}$ are closed (in the standard topology on \mathcal{A}^H).

A4. SEPARABILITY. For every p, H and $x, y, z, w \in \mathcal{A}^H$, if $(x+z)*(y+w) = 0$, $x \geq_{p,H} y$, and $z \geq_{p,H} w$, then $(x+z) \geq_{p,H} (y+w)$.

A2 is a standard monotonicity axiom. It will turn out to imply that the similarity function is nonnegative. Without it one may

obtain a numerical representation as in (*) where the similarity function is not constrained in sign. A3 is a continuity axiom. It guarantees that, if $x_k \geq_{p,H} y$ and $x_k \rightarrow x$, then $x \geq_{p,H} y$ also holds (and similarly $x_k \leq_{p,H} y$ implies that $x \leq_{p,H} y$).

From a conceptual viewpoint, the separability axiom A4 is our main assumption. It states that preferences can be “added up.” That is, if two act profiles, x and z , are (weakly) preferred to two others, y and w , respectively, then the sum of the former is (weakly) preferred to the sum of the latter, provided that such preferences are well defined. It is powerful enough to preclude cyclical strict preferences. Moreover, A4 will play a crucial role in showing that the numerical representation is additive across cases, as well as that the effect of each past case may be represented by the product of the utility of the result and the similarity of the problem.

We do not attempt to defend A4 as “universally reasonable.” On the contrary, we readily agree that it may be too restrictive for some purposes.³ For instance, one may certainly consider an additive functional with a case-dependent utility, as in theories of state-dependent expected utility theory, or a nonseparable functional. Alternatively, one may allow the similarity function to be modified according to the results that the agent has experienced. For the time being we merely offer an axiomatization of a case-based decision theory, which may be viewed as a “first approximation.” The main role of the axioms above is not to convince the reader that our theory is reasonable. Rather, our main goal is to show that the theoretical concept of “similarity,” combined with U -maximization, is in principle derivable from observed preferences.

The first result can finally be presented.

THEOREM 1. The following two statements are equivalent:

- (i) A1–A4 hold;
- (ii) For every $p \in P$ and every H there exists a function

$$s_{p,H} : H \rightarrow \mathcal{R}_+$$

$$x \geq_{p,H} y \quad \text{iff} \quad \sum_{q \in H} s_{p,H}(q)x(q) \geq \sum_{q \in H} s_{p,H}(q)y(q)$$

for all compatible $x, y \in \mathcal{R}^H$.

3. Note, however, that A4 may appear very restrictive partly because of our simplifying assumption that results are represented by utilities.

Furthermore, in this case, for every p, H , the function $s_{p,H}$ is unique up to multiplication by a positive scalar.

Setting $s(p, q) = s_{p,H}(q)$, Theorem 1 gives rise to U -maximization for a given set of problems H . That is, considering the actual memory M the agent possesses at the time of decision p , she would choose an act that maximizes the formula (*) with $s(p, q) = s_{p,H}(q)$ and $H = H(M)$. However, this similarity function may depend on the set of problems H . The next axiom ensures that the similarity measure is independent of memory. Specifically, A5 compares the relative importance of two problems, q_1 and q_2 , in two histories, H^1 and H^2 . It requires that the similarity weights assigned to these problems in the two histories be proportional.

A5. SIMILARITY INVARIANCE. For every $p, q_1, q_2 \in P$ and every two memories M^1, M^2 with $q_1, q_2 \in H^i \equiv H(M^i)$ ($i = 1, 2$) and $p \notin H^i$ ($i = 1, 2$), let v_j^i stand for the unit vector in \mathcal{R}^{H^i} ($i = 1, 2$) corresponding to q_j ($j = 1, 2$). (That is, v_j^i is a vector whose q_j th component is 1 and its other components are 0.) Then, denoting the symmetric part of $\geq_{p,H}$ by $\approx_{p,H}$,

$$x, y \in \mathcal{R}^{H^1}, z, w \in \mathcal{R}^{H^2}, x \approx_{p,H^1} y, z \approx_{p,H^2} w$$

and

$$x + \alpha v_1^1 \approx_{p,H^1} y + \beta v_2^1$$

imply that

$$z + \alpha v_1^2 \approx_{p,H^2} w + \beta v_2^2$$

whenever the compared profiles are compatible.

Equipped with A5, one may define a single similarity function that represents preferences given any history.

THEOREM 2. The following two statements are equivalent:

- (i) A1–A5 hold.
- (ii) There exists a function $s: P^2 \rightarrow [0, 1]$ such that for all $p \in P$, every memory M with $p \notin H = H(M)$ and every compatible $x, y \in \mathcal{R}^H$,

$$x \geq_{p,H} y \quad \text{iff} \quad \sum_{q \in H} s(p, q)x(q) \geq \sum_{q \in H} s(p, q)y(q).$$

Furthermore, in this case, for every p , the function $s(p, \cdot)$ is unique up to multiplication by a positive scalar.

Note that the decision rule axiomatized here is U -maximization as discussed in the introduction. The proofs of both theorems are given in Appendix 1.

III. DISCUSSION

III.1. The Model

Subjective Similarity. The similarity function in our model is derived from preferences, and is thus "subjective." That is, different individuals will typically have different preferences, which may give rise to different similarity functions, just as preferences give rise to subjective probability in the works of de Finetti [1937] and Savage [1954]. Yet, for some applications one may wish to have a notion of "objective similarity," comparable to "objective probability."

Anscombe and Aumann [1963] define objective probability as a nickname for a subjective probability measure, which happens to be shared by all individuals involved. By a similar token, if the subjective similarity functions of all relevant agents happen to coincide, we might dub this common function objective similarity. Alternatively, one may argue that objectivity of a certain cognitive construct—such as probability or similarity—entails more than a mere (and perhaps coincidental) identity of its subjective counterpart across individuals. Indeed, some feel that objectivity requires some justification. Be that as it may, objective similarity is in particular also the subjective similarity of those individuals who accept it.

For purposes of objective similarity judgments, as well as for normative applications, our similarity function may be too permissive. For instance, we have not required it to be symmetric. One may wonder under what conditions can the similarity function $s(p, \cdot)$ of Theorem 2 be rescaled (separately for each p) so that $s(p, q) = s(q, p)$ for all $p, q \in P$. It turns out that a necessary and sufficient condition is that for all $p, q, r \in P$,

$$s(p, q)s(q, r)s(r, p) = s(p, r)s(r, q)s(q, p).$$

(Note that this condition does not depend on the choice of $s(p, \cdot)$, $s(q, \cdot)$, and $s(r, \cdot)$.)⁴ However, in view of psychological evidence

4. This condition can be translated to original data, namely, to observed preferences. Such a formulation will be more cumbersome without offering any theoretical advantage. Since the similarity functions are derived from preferences in an essentially unique manner, we may use them in the formulation of additional

[Tversky 1977], this can be unduly restrictive for a descriptive theory of subjective similarity.

Other Interpretations. In the development of CBDT we advance a certain cognitive interpretation of the functions u and s . However, the theory can also accommodate alternative, behaviorally equivalent interpretations. First, consider the function u . We assumed that it represents fixed preferences, and that memory may affect choices only by providing information about the u -value that certain acts yielded in the past. Alternatively, one may suggest that memory has a direct effect on preferences. According to this interpretation, the utility function is the aggregate U , while the function u describes the way in which U changes with experience. For instance, if the agent has a high aspiration level—corresponding to negative u values—she will like an option less, the more she used it in the past, and will exhibit change-seeking behavior. On the other hand, a low aspiration level—positive u values—would make her “happier” with an option, the more she is familiar with it, and would result in habit formation. In Gilboa and Schmeidler [1993] we develop a model of consumer choices based on this interpretation.

The function s can also have more than one cognitive interpretation. Specifically, when the agent is faced with a decision problem, she may not recall all relevant cases. The probability that a case be recalled may depend on its salience, the time that elapsed since it was encountered, and so forth. Thus, our function s should probably be viewed as reflecting both probability of recall and “intrinsic” similarity judgments.⁵

When “behavior” is understood to mean “revealed preference” (as opposed to, say, speech), one probably cannot hope to disentangle various cognitive interpretations based on behavioral data. Whereas specific applications may favor one interpretation over another, predictions of behavior would not depend on the cognitive interpretation chosen.

Hypothetical Cases. Consider the following example. An agent has to drive to the airport in one of two ways. When she gets there safely, she learns that the other road was closed for construction. A

axioms without compromising the behavioral content of the latter. Indeed, A5 could also be more elegantly formulated in terms of the derived similarity functions.

5. Some readers expressed preference for the terms “relevance” or “weight” over “similarity.” Others insisted that we should use “payoff” rather than “utility.” We find these alternative terms completely acceptable.

week later she is faced with the same problem. Regardless of her aspiration level, it seems obvious that she will choose the same road again. (Road constructions, at least in psychologically plausible models, never end.)

Thus, relevant cases may also be hypothetical, or counterfactual. ("If I had taken the other way, I would never have made it.") Hypothetical cases may endow a case-based decision-maker with reasoning abilities she would otherwise lack. It seems that any knowledge the agent possesses and any conclusions she deduces from it can, inasmuch as they are relevant to the decision at hand, be reflected by hypothetical cases.

Average Performance. The functional U gathers data in an additive way. For instance, assuming that all problems are equally similar, an act that was tried ten times with a u -value of 1 will be ranked higher than an act that was tried only once and resulted in a u -value of 5. One may therefore be interested in a decision rule that maximizes the following functional:

$$(**) \quad V(a) = \sum_{(q,a,r) \in M} s'(p,q)u(r),$$

where

$$s'(p,q) = \left\{ \begin{array}{ll} \frac{s(p,q)}{\sum_{(q',a,r) \in M} s(p,q')} & \text{if well defined} \\ 0 & \text{otherwise} \end{array} \right\}$$

and $s(p,q)$ is the similarity function of Section II. According to this formula, for every act a the similarity coefficients $s'(p,q)$ add up to one (or to zero). Note that this similarity function depends not only on the problem encountered in the past, but also on the acts chosen at different problems.

Observe that V is discontinuous in the similarity values at zero. For example, if an act a was chosen in a single problem q and resulted in a very desirable outcome, it will have a high V -value as long as $s(p,q) > 0$ but will be considered a "new act," with zero V -value, if $s(p,q) = 0$. In the Bosnia example, for instance, V maximization may lead to different decisions depending on whether the Gulf War is considered to be "remotely relevant" or "completely irrelevant."⁶ By contrast, the functional U is continuous in the similarity values.

6. Observe, however, that discontinuity can only occur if all past cases are at most remotely relevant.

Of special interest is the case where $s(p,q) = 1$ for all $p,q \in P$. In this case, V is simply the average utility of each act. The condition $s(\cdot, \cdot) = 1$ means that, at least as far as the agent's preferences reveal, all problems are basically identical. In this case, this variant of case-based decision theory is equivalent to "frequentist expected utility theory": the agent chooses an act with maximal "expected" utility, where the outcome distribution for each act is assumed to be given by the observed frequencies. (Note also that in this particular model the discontinuity at $s(\cdot, \cdot) = 0$ does not pose a problem, since $s(\cdot, \cdot) \equiv 1$.) In Appendix 2 we provide an axiomatization of V -maximization.

The Definition of Acts. Case-based decision-makers may appear to be extremely conservative and boring creatures: once an act achieves their aspiration level, they stick to it. Our agent, it would seem, is an animal that always eats the same food at the same place, chooses the same form of entertainment (if at all), and so forth.

Although this is true at some level of description, it does not have to be literally true. For instance, the act that is chosen over and over again need not be "Have lunch at X"; it may also be "Have lunch at a place I did not visit this week." Repetition at this level of description will obviously generate an extremely diverse lunch pattern.

III.2 The Axiomatization

Observability of Preferences and Hypothetical Questions. Whenever we encounter our agent, she has a certain memory M and can only exhibit preferences complying with $\geq_{p,M}$. It is therefore natural to ask, in what sense is the relation $\geq_{p,H}$ observable?

An experimenter may try to access the agent's preferences for different memories by confronting her with (i) counterfactual choices among acts, or (ii) actual choices among "strategies." In the first case, the agent may be asked to rank acts not only based on their *actual* act profiles, but also based on act profiles they *may have had*. Thus, she may be asked, "Assume act a yielded r in problem q . Would you still prefer it to act b ?" In the second case, the agent may only be given the set of problems H , and then be asked to choose a strategy, that is, to make her choice contingent upon the act profiles which were not revealed to her.

In both procedures, one may distinguish between two levels of

hypothetical (or conditional) questions. Suppose that the agent prefers act a to b , and consider the following types of questions.

- I. Remember the case $c = (p, a, r)$, where you chose a and got r ? Well, assume that the outcome were t instead of r . Would you still prefer a to b ?
- II. Remember the case $c = (p, a, r)$ where you chose a and got r ? Well, now imagine you actually chose another act a' and received t . Would you still prefer a to b ? How about a' to b ?

In Section II we implicitly assumed that questions of both types can be meaningfully answered. Yet one may argue that questions of type II are too hypothetical to serve as foundations of any behavioral decision theory. While the agent has no control over the outcome r , she may insist that in problem p she would never have tried act a' and that the preference question is meaningless.

Appendix 2 presents a model in which answers to questions only of the first type are assumed. We provide an axiomatic derivation of the linear evaluation functional with a similarity function which, unlike that in Theorem 1, depends not only on the problems encountered, H , but also on the actions that were chosen in each. This more general functional form allows us to axiomatize V -maximization as a special case.

Derivation of Utility. The axiomatization provided here presupposes that the set of results is \mathcal{R} , and that results are measured in utiles, namely, that the utility function is linear. Thus, our axiomatic derivation of the notion of similarity and the CBDT functional relies on a supposedly given notion of utility, in a manner that parallels de Finetti's [1937] axiomatization of subjective probability together with expected utility maximization. Needless to say, the concept of a utility function is also a theoretical construct that calls for an axiomatic derivation from observable data. Ideally, one would like to start out with a model that presupposes neither similarity nor utility, and to derive them simultaneously, in conjunction with the CBDT decision rule. Such a derivation would also highlight the fact that the utility function, like the similarity function in Theorem 1, may, in general, depend on p, H . However, to keep the axiomatization simple, we do not follow this track here.

IV. CBDT AND EUT

Complementary Theories. We do not consider case-based decision theory "better" than or as a substitute for expected utility

theory. Rather, we view them as complementary theories. The classical derivation of EUT, as well as the derivation of CBDT in this paper, are behavioral in that the theoretical constructs in these models are induced by observable (in principle) choices. Yet the scope of applicability of these theories may be more accurately delineated if we attempt to judge the psychological plausibility of the various constructs. Two related criteria for classification of decision problems may be relevant. One is the problem's description; the second is its relative novelty.

If a problem is formulated in terms of probabilities, for instance, EUT is certainly a natural choice for analysis and prediction. Similarly, when states of the world are naturally defined, it is likely that they would be used in the decision-maker's reasoning process, even if a (single, additive) prior cannot be easily formed. However, when neither probabilities nor states of the world are salient (or easily accessible) features of the problem, CBDT may be more plausible than EUT.

We may thus refine Knight's [1921] distinction between risk and uncertainty by introducing a third category of "ignorance": risk refers to situations where probabilities are given; uncertainty to situations in which states are naturally defined, or can be simply constructed, but probabilities are not. Finally, decision under ignorance refers to decision problems for which states are neither (i) "naturally given" in the problem nor (ii) can they be easily constructed by the decision-maker. EUT is appropriate for decision-making under risk. In the face of uncertainty (and in the absence of a subjective prior) one may still use those generalizations of EUT that were developed to deal with this problem specifically, such as nonadditive probabilities [Schmeidler 1989] and multiple-priors [Bewley 1986; Gilboa and Schmeidler 1989]. However, in cases of ignorance, CBDT is a viable alternative to the EUT paradigm.

Classifying problems based on their novelty, one may consider three categories. We suggest that CBDT is useful at the extremes of the novelty scale, and EUT in the middle. When a problem is repeated frequently enough, such as whether to stop at a red traffic light, the decision becomes almost automated and "rule-based." Such decisions may be viewed as a special type of case-based decisions. Indeed, a rule can be thought of as a summary of many cases, from which it was probably derived in the first place.⁷ When

7. See the discussion of "ossified cases" in Riesbeck and Schank [1989] and of induction and rules in Gilboa and Schmeidler [1993b].

deliberation is required, but the problem is familiar, such as whether to buy insurance, it can be analyzed "in isolation": its own history suffices for the formulation of states of the world and perhaps even a prior, and EUT (or some generalization thereof) may be cognitively plausible. Finally, if the problem is unfamiliar, such as whether to get married or to invest in a politically unstable country, it needs to be analyzed in a context- or memory-dependent fashion, and CBDT is again a more accurate description of the way decisions are made.

Reduction of Theories. While CBDT may be a more natural framework in which to model satisficing behavior, EUT can be used to explain this behavior as well. For instance, the Bayesian-optimal solution to the famous "multi-armed bandit" problem [Gittins 1979] may not ever attempt to choose certain options. In fact, it is probably possible to provide an EUT account of any application in which CBDT can be used, by using a rich enough state space and an elaborate enough prior on it. Conversely, one may also "simulate" an expected utility maximizer by a case-based decision-maker whose memory contains a sufficiently rich set of hypothetical cases: given a set of states of the world Ω and a set of consequences R , let the set of acts be $A = R^\Omega = \{a: \Omega \rightarrow R\}$. Assume that the agent has a utility function $u: R \rightarrow \mathcal{R}$ and a probability measure μ on Ω . (Where Ω is a measurable space. For simplicity, it may be assumed finite.) The corresponding case-based decision-maker would have a hypothetical case for each pair of a state of the world ω and an act a :

$$M = \{((\omega, a), a, a(\omega)) \mid \omega \in \Omega, a \in A\}.$$

Letting the similarity of the problem at hand to the problem (ω, a) be $\mu(\omega)$, U -maximization reduces to expected utility maximization. (Naturally, if Ω or R are infinite, one would have to extend CBDT to deal with an infinite memory.) Furthermore, Bayes' update of the probability measure may also be reflected in the similarity function: a problem whose description indicates that an event $B \subseteq \Omega$ has occurred should be set similar to degree zero to any hypothetical problem (ω, a) , where $\omega \notin B$.

Since one can mathematically embed CBDT in EUT and vice versa, it is probably impossible to choose between the two on the basis of predicted observable behavior.⁸ Each is a refutable theory given a description of a decision problem, where its axioms set the

8. Matsui [1993] formally proves an equivalence result between EUT and CBDT. His construction does not resort to hypothetical cases.

conditions for refutation. But in most applications there is enough freedom in the definition of states or cases, probability or similarity, for each theory to account for the data. Moreover, a problem that is formulated in terms of states has many potential translations to the language of cases and vice versa. It is therefore hard to imagine a clear-cut test that will select the "correct" theory.

To a large extent, EUT and CBDT are not competing theories; they are different *languages*, in which specific theories are formulated. Rather than asking which one of them is more accurate, we should ask which one is more convenient. The two languages are equally powerful in terms of the range of phenomena they can describe. But for each phenomenon, they will not necessarily be equally intuitive. Furthermore, the specific theories we develop in these languages need not provide the same predictions given the same observations. Hence we believe that there is room for both languages.

Asymptotic Behavior. One may wonder whether, when the same problem is repeated over and over again, CBDT would converge to the choice prescribed by EUT for the one-shot problem with known probabilities. This does not appear to be the case if we take CBDT to mean either *U*- or *V*-maximization.

Consider the following setup: $A = \{a, b\}$, $s(\cdot, \cdot) \equiv 1$. Assume that nature chooses the outcomes for each act by given distributions in an independent fashion. That is, there are two random variables R_a, R_b such that whenever the agent chooses a (b), the outcome is chosen according to a realization of R_a (R_b), independently of past choices and realizations. Further, assume the following distributions:

$$R_a = \begin{pmatrix} 1 & 0.6 \\ -1 & 0.4 \end{pmatrix}; \quad R_b = \begin{pmatrix} 100 & 0.7 \\ -2 & 0.3 \end{pmatrix}.$$

First consider a *U*-maximizer agent. At the beginning, both a and b have identical (empty) histories, and the decision is arbitrary. Suppose that the agent chooses a with probability .5, and that R_a results in +1. From then on she will choose a as long as the random walk generated by these choices is positive. Hence there is a positive probability that she will always choose a . Next consider a *V*-maximizer agent. Suppose that she first chose b , and that it resulted in the outcome -2. From that stage on this agent will always choose a .

Thus, for both decision rules we find that there is a positive

probability that the agent will not maximize the "real" expected utility even in cases where objective probabilities are defined. Arguably, it is in these cases that EUT is most appealing. However, in Gilboa and Schmeidler [forthcoming] we show that, if the aspiration level is adjusted in an appropriate manner over time, U -maximization will converge to expected u maximization in the long run.⁹

Hypothetical Reasoning. Judging the cognitive plausibility of EUT and CBDT, one notes a crucial difference between them: CBDT, as opposed to EUT, does not require the decision-maker to think in hypothetical or counterfactual terms. In EUT, whether explicitly or implicitly, the decision-maker considers states of the world and reasons in propositions of the form, "If the state of the world were ω and I chose a , then r would result." In CBDT no such hypothetical reasoning is assumed.

Similarly, there is a difference between EUT and CBDT in terms of the informational requirements they entail regarding the utility function: to "implement" EUT, one needs to know the utility function u , i.e., its values for any consequence that may result from any act. For CBDT, on the other hand, it suffices to know the u -values of those outcomes that were actually experienced.

The reader will recall, however, that our axiomatic derivation of CBDT involved preferences among hypothetical act profiles. It might appear therefore that CBDT is no less dependent on hypothetical reasoning than EUT. But this conclusion would be misleading. First, one has to distinguish between elicitation of parameters by an outside observer, and application of the theory by the agent herself. While the elicitation of parameters such as the agent's similarity function may involve hypothetical questions, a decision-maker who knows her own tastes and similarity judgments need not engage in any hypothetical reasoning in order to apply CBDT. By contrast, hypothetical questions are intrinsic to the application of EUT.

Second, when states of the world are not naturally given, the elicitation of beliefs for EUT also involves inherently hypothetical questions. Classical EUT maintains that no loss of generality is involved in assuming that the states of the world are known, since

9. It was pointed out to us by Avraham Beja that, should one adapt our model to derive the utility function u axiomatically, the latter may or may not coincide with von-Neumann-Morgenstern utility function derived from choices among lotteries.

one may always define the states of the world to be all the functions from available acts to conceivable outcomes. This view is theoretically very appealing, but it undermines the supposedly behavioral foundations of Savage's model. In such a construction, the set of "conceivable acts" one obtains is much larger than the set of acts from which the agent can actually choose. Specifically, let there be given a set of acts A and a set of outcomes X . The states of the world are X^A , i.e., the functions from acts to outcomes. The set of conceivable acts will be $\bar{A} \equiv X^{(X^A)}$, that is, all functions from states of the world to outcomes. Hence the cardinality of the set of conceivable acts \bar{A} is by two orders of magnitude larger than that of the actual ones A . Yet, using a model such as Savage's, one needs to assume a (complete) preference order on \bar{A} , while *in principle* preferences can be observed only between elements of A . Differently put, such a "canonical construction" of the states of the world gives rise to preferences that are intrinsically hypothetical, and is a far cry from the behavioral foundations of Savage's original model.

In summary, in these problems both EUT and CBDT rely on hypothetical questions or on "contingency plans" for elicitation of parameters. The Savage questionnaire to elicit EUT parameters will typically involve a much larger and less intuitive set of acts than the corresponding one for CBDT. Furthermore, when it comes to application of the theory, CBDT clearly requires less hypothetical reasoning than EUT.

Cognitive and Behavioral Validity. CBDT may reflect the way people think about certain decision problems better than EUT. But many economists would argue that we should not care how agents think, as long as we know how they behave. Moreover, they would say, Savage's behavioral axioms are very reasonable; thus, it is very reasonable that people would behave *as if* they were expected utility maximizers. However, we claim that behavioral axioms which appear plausible assuming the EUT models are not as convincing when this very model is unnatural. For instance, Savage's "sure-thing principle" (his axiom P2) is very compelling when acts are given as functions from states to outcomes. But in examples such as 2 and 3 in the Introduction, outcomes and states are not given, and it is not clear what all the implications of the sure-thing principle are. It may even be hard to come up with an example of acts that are actually available in these examples, and such that the sure-thing principle constrains preferences among

them. It is therefore not at all obvious that actual behavior would follow this seemingly very compelling principle. More generally, the predictive validity of behavioral axioms is not divorced from the cognitive plausibility of the language in which they are formulated.

V. APPLICATIONS

This section is devoted to economic applications of case-based decision theory. All we could hope to provide here are some sketchy illustrations, which certainly fall short of complete models. Our goal is merely to suggest that CBDT may be able to explain some phenomena in a simpler and more intuitive way than EUT.

V.1. *To Buy or Not to Buy*

Consider the following example. A firm is about to introduce two new products $\{1,2\}$ into a market. When product i is introduced, the consumers face a decision problem p_i , with two possible acts $\{a,b\}$, where b stands for buying the product and a for abstaining from purchase. A consumer's decision to buy product i , say, a cereal or a soup, implies consumption on a regular basis in given quantities. The consumers are familiar with product 0 of the same firm. Product 1 is similar to both products 0 and 2, but the latter are not similar to each other. Finally, each consumer will derive a positive utility level from each product consumed.

In this case, the order in which the products are introduced may make a difference. If the firm introduces product 1 and then product 2, both will be purchased. However, if product 2 is introduced first, a consumer's memory contains nothing that resembles it at the time of decision. Thus, her choice between a and b will be arbitrary, and she may decide not to buy the product. As a result, we expect a lower aggregate demand for product 2 if it is introduced first than if it is introduced after product 1.

While EUT-based models could also provide such behavioral predictions, we find it more plausible that consumer decisions are directly affected by perceived similarities. Indeed, advertising techniques often seem to exploit and even manipulate the consumer's similarity judgments.

V.2. *Reputation*

Case-based consumer decisions give rise to aspects of reputation quite naturally. Consider a model with two products and two firms. Assume that product 1 is produced only by firm A. Product 2

is new. It is produced by both firms *A* and *B*. Other things being equal, firm *A* will have an edge in market 2 if it satisfies consumers' expectations in market 1 (i.e., if $U(A) > 0$). Thus, one would expect successful firms to enter new markets even if the technology needed in them is completely different from that used in the traditional ones.

An EUT explanation of the role of reputation would typically involve consumers' beliefs about the firms' rationality, as well as beliefs about the firms' beliefs about consumers. CBDT makes much weaker rationality assumptions in explaining this phenomenon.

V.3. Introductory Offers

Another phenomenon that is close in nature is the introduction of new products at discounted rates. Again, one may explain the optimality of such marketing policies with "fully rational" expected utility consumers. For instance, in the presence of experimentation cost or risk aversion, a fully rational consumer may tend to buy the product at the regular price after having bought it at the introductory (lower) price. Yet if consumers are case-based decision-makers, the formation of habits is a natural feature of the model.

VI. VARIATIONS AND FURTHER RESEARCH

Memory-Dependent Similarity. In reality, similarity judgments may depend on the results obtained in past cases. For instance, the agent may realize that certain attributes of a problem are more or less important than she previously believed. In Gilboa and Schmeidler [1994b] we dub this phenomenon "second-order induction," and discuss the relationship between CBDT and the process of induction in more detail.

When similarity is memory-dependent, two assumptions of our model may be violated. First, the separability axiom A4 may fail to hold. Second, the assumption that acts are ranked based on their act profiles may also be too restrictive. Specifically, the choice among acts need not satisfy independence of irrelevant alternatives, and it therefore cannot be represented by a binary relation over act profiles. Thus, CBDT in its present form does not describe how agents learn the similarity function.

Similar Acts. In certain situations, an agent may have some information regarding an act without having tried it in the past.

For instance, the agent may consider buying a house in a neighborhood where she has owned a house before. The experience she had with a different, but similar, act is likely to color her evaluation of the one now available to her. Furthermore, some acts may involve a numerical parameter, such as "Offer to sell at a price p ." One would expect the evaluation of such acts to depend on past performance of similar acts with a slightly different value of the parameter.

These examples suggest the following generalization of CBDT: consider a similarity function over (problem,act) pairs; given a certain memory and a decision problem, every act is compared—in conjunction with the current problem—to *all* (problem,act) pairs in memory, and a similarity weighted utility value is computed for it. Maximization of such a function is axiomatized in Gilboa and Schmeidler [1994a].

Act Generation. It is often the case that the set of available acts is not naturally given and has to be constructed by the agent. CBDT as presented here is not designed to deal with these problems, and it may certainly benefit from insights into the process of "act generation." In particular, the vast literature on planning in artificial intelligence may prove relevant to modeling of decision-making under uncertainty.

Changing Utility. The framework used in Section II, in which outcomes are identified with utility levels, is rather convenient to convey the main idea, but it may also be misleading: it entails the implicit assumption that the utility function does not depend on the memory M , on time (which may be implicit in M), and so forth.

There may be some interest in a more general model, where the utility is allowed to vary with memory. In particular, the utility scale may "shift" depending on one's experience. Recall that the utility is normalized so as to set $u(r_0) = 0$. As mentioned above, one may refer to this value as the aspiration level of the decision-maker. A shift of the utility function is therefore equivalent to a change in the agent's aspiration level.

Adopting this cognitive interpretation, it is indeed natural that the aspiration level be adjusted according to past achievements. In Gilboa and Schmeidler [1994a] we axiomatize a family of decision rules that allow the aspiration level to be a linear function of the outcomes experienced in the past. However, some applications may resort to nonlinear adjustment rules as well. (See, for instance,

Gilboa and Schmeidler [forthcoming].) The axiomatic foundations of aspiration level adjustments therefore call for further research.

Normative Interpretation. Our focus in this paper is on CBDT as a descriptive theory, which, for certain applications, may be more successful than EUT. Yet, in some cases CBDT may also be a more useful normative theory. While we share the view that it is desirable and "more rational" to think about all possible scenarios and reason about them in a consistent way, we also hold that a normative theory should be practical. For instance, if the state space is huge, and the agent does not entertain probabilistic beliefs over it, telling her that she *ought* to have a prior may be of little help.

If we believe that, in a given problem, applying EUT is not a viable option, we might at least attempt to improve case-based decisions. For instance, one may try to change one's similarity function so that it be symmetric, ignore primacy and recency effects (i.e., resist the tendency to assign disproportionate similarity weights to the first and the most recent cases), and so forth. It might even be argued that it is more useful to train professionals (doctors, managers, etc.) to make efficient and probably less biased case-based decisions rather than to teach them expected utility theory. However, such claims and the research that is needed to support them are beyond the scope of this paper.

Welfare Implications. A cognitive interpretation of CBDT raises some welfare questions. Is a satisfied individual "happier" than an unsatisfied one? Should the former be treated as richer simply because she has a lower aspiration level? Should we strive to increase people's aspiration levels, thereby prodding them to perform better? Or should we lower expectations so that they are content? We do not dwell on these questions here, partly because we have no answers to offer.

Strategic Aspects. In a more general model, one may try to capture manipulations of the similarity function. In phenomena as diverse as advertising and legal procedures, people try to influence other peoples' perceived similarity of cases. Moreover, an agent may wish to expose other agents to information selectively, in a way that will bring about certain modes of behavior on their part.

Procedures of Recall. CBDT may greatly benefit from additional psychological insights into the structure of memory and from empirical findings regarding the recollection process. For

instance, one may hypothesize that the satisficing nature of decision-making is revealed not only in a dynamic context, but also within each decision: rather than computing the U -value of *all* possible acts, the agent may stop at the first act which obtains a positive U -value. There are, however, several ways in which "first" could be defined. For example, the agent may ask herself, "When did I choose this act?" and only after the evaluation of a given act will the next one be considered. Alternatively, she may focus on the problem and ask, "When was I in a similar situation?" and as the cases are retrieved from memory one by one, the function U is updated for all acts—until one act exceeds the aspiration level. These two models induce different decision rules.

Similarly, insight may be gained from analyzing the structure of a "decision problem" and the corresponding structure of the similarity function in specific contexts. Some psychological studies relating to this problem are Gick and Holyoak [1980, 1983] and Falkenhainer, Forbus, and Gentner [1989].

Other Directions. The model we present here should be taken merely as a first approximation. Just as EUT encountered the "paradoxes" of Allais [1953] and of Ellsberg [1961], the linear functional we propose here is likely to be found too restrictive in similar examples. Correspondingly, almost every generalization of EUT may have a reasonable counterpart for CBDT.

The main goal of this paper was to explore the possibility of a formal, axiomatically based decision theory, using a less idealized and at times more realistic paradigm than EUT. We believe that case-based decision theory is such an alternative.

APPENDIX 1: PROOF OF THEOREMS

Regarding both theorems, the fact that the axioms are necessary for the desired representation is straightforward. Similarly, the uniqueness of the similarity functions is simple to verify. We therefore provide here only proofs of sufficiency, that is, that our axioms imply the numerical representations.

Proof of Theorem 1. Fix p, H , and denote $\geq \cdot = \geq_{p,H}$. We also use the notation $X = \mathcal{P}^H$ and identify it with \mathcal{P}^n for $n = |H|$. W.l.o.g. assume that $H \neq \emptyset$. First, note the following.

Observation If $\geq \cdot$ satisfies A1 and A4, then

- (i) for all $x, y \in X$ with $x * y = 0$, $x \geq \cdot y \Leftrightarrow -y \geq \cdot -x$;

- (ii) for all $x, y, z, w \in X$ with $x * y = 0$, $(x + z) * (y + w) = 0$ and $z \approx w$, $x \geq y \Leftrightarrow (x + z) \geq (y + w)$.

Proof. (i) Assume that $x \geq y$. Consider $z = w = -(x + y)$, and use A4 (where $z \geq w$ follows from A1).

- (ii) Under the provisions of the claim, $z \geq w$, and A4 implies that $x \geq y \Rightarrow (x + z) \geq (y + w)$.

As for the converse, define $z' = -z$ and $w' = -w$. By (i), $-z \geq -w$. Thus, A4 can be used again to conclude $x \geq y$.

We now turn to the proof of Theorem 1. Define $\geq' \subseteq X \times X$ by

$$x \geq' y \Leftrightarrow (x - y) \geq 0 \quad \text{for all } x, y \in \mathcal{R}^n = X.$$

We need several lemmata whose proofs are rather simple. For brevity's sake we merely indicate which axioms and lemmata are used in each, omitting the details:

LEMMA 1. For $x, y \in X$ with $x * y = 0$, $x \geq' y$ iff $x \geq y$ (the observation above).

LEMMA 2. \geq' is complete, i.e., for all $x, y \in X$, $x \geq' y$, or $y \geq' x$ (the definition of \geq' and A1).

LEMMA 3. \geq' is transitive (the definition of \geq' and A4).

LEMMA 4. \geq' is monotone, i.e., for all $x, y \in X$, $x \geq y$ implies that $x \geq' y$ (the definition of \geq' and A2).

LEMMA 5. \geq' is continuous, i.e., for all $x \in X$ the sets $\{y \in X \mid y \geq' x\}$, $\{y \in X \mid x \geq' y\}$ are closed in \mathcal{R}^n (the definition of \geq' and A3). (In view of Lemma 2, this is equivalent to the sets $\{y \in X \mid y >' x\}$, $\{y \in X \mid x >' y\}$ being open.)

LEMMA 6. \geq' satisfies the following separability condition: for all $x, y, z \in X$, $x \geq' y$ if and only if $(x + z) \geq' (y + z)$ (the definition of \geq').

LEMMA 7. \geq' satisfies the following condition: for all $x, y, z, w \in X$, if $x \geq' y$ and $z \geq' w$, then $(x + z) \geq' (y + w)$ (the definition of \geq' and A4).

LEMMA 8. If $x \geq' y$, then $x \geq' (x + y)/2 \geq' y$ (Lemmata 2, 3, and 6).

LEMMA 9. If $x \geq' y$ and $\alpha \in (0, 1)$, then $x \geq' \alpha x + (1 - \alpha)y \geq' y$ (successive application of Lemma 8, in conjunction with Lemmata 3 and 5).

LEMMA 10. For every $x \in X$, the sets $\{y \in X | y \geq' x\}$, $\{y \in X | x \geq' y\}$, $\{y \in X | y >' x\}$, and $\{y \in X | x >' y\}$ are convex. (Lemmata 3 and 9).

LEMMA 11. Define $A = \{x \in X | x \geq' 0\}$ and $B = \{x \in X | 0 >' x\}$ (where, as above, 0 denotes the zero vector in $X = \mathcal{R}^n$). Then A is nonempty, closed and convex; B is open and convex; $A \cap B = \emptyset$; and $A \cup B = \mathcal{R}^n$ (Lemmata 2, 5, and 10).

LEMMA 12. If $B = \emptyset$, the function $s(\cdot) \equiv 0$ satisfies the representation condition. If $B \neq \emptyset$, there exist a nonzero linear functional $S: \mathcal{R}^n \rightarrow \mathcal{R}$ and a number $c \in \mathcal{R}$ such that

$$\begin{aligned} S(x) &\geq c \text{ for all } x \in A \\ S(x) &< c \text{ for all } x \in B \end{aligned}$$

(in view of Lemma 11, a standard separating-plane argument).

LEMMA 13. In the case $B \neq \emptyset$, the constant c in Lemma 12 is zero, hence $S(x) \geq 0$ if and only if $x \in A$ ($c \leq 0$ follows from Lemma 2; $c \geq 0$ is a result of Lemma 7).

By Lemmata 1 and 7, for every compatible $x, y \in X$, $x \geq' y$ iff $(x - y) \geq' 0$, i.e., iff $(x - y) \in A$. If $B = \emptyset$, Lemma 12 concludes the proof. If $B \neq \emptyset$, the function $s: H \rightarrow \mathcal{R}$ defined by S satisfies the desired representation by Lemma 13. Furthermore, it is nonnegative by Lemma 4.

Remark. Note that we have also proved that A1, A3, and A4 are necessary and sufficient for a numerical representation as in (*), where the similarity function s is not restricted to be nonnegative.

Proof of Theorem 2. Theorem 1 guarantees that for every $p \in P$ and $H \subseteq P$ with $p \notin H$ there exists a function $s_H(p, \cdot) \equiv s_{p,H}(\cdot): H \rightarrow \mathcal{R}_+$ such that

$$x \geq_{p,H} y \text{ iff } \sum_{q \in H} s_H(p,q)x(q) \geq \sum_{q \in H} s_H(p,q)y(q)$$

for every compatible $x, y \in \mathcal{R}^H$. We wish to show that for every $p \in P$ there is a *single* function $s(p, \cdot)$ satisfying the condition above for every history $H \subseteq P \setminus \{p\}$.

Theorem 1 also states that each of the functions $s_H(p, \cdot)$ is unique, but only up to a positive multiplicative scalar. Thus, it

suffices to show that for every $p \in P$ there exists a function $s(p, \cdot)$, such that for every $H \subseteq P \setminus \{p\}$ there exists a coefficient $\lambda_{p,H} > 0$ such that

$$s(p, q) = \lambda_{p,H} s_H(p, q) \quad \text{for all } q \in H.$$

Fix a problem $p \in P$. We first define the function $s(p, \cdot)$, and will then show that there are coefficients $\lambda_{p,H} > 0$ as required. Set $H^0 = P \setminus \{p\}$. Since P is finite, so is H^0 . We set

$$s(p, q) = s_{H^0}(p, q) \quad \text{for all } q \neq p.$$

(One may generalize our results to an infinite set of problems, out of which only finitely many may appear in any history. In this case one may not use a maximal finite $H \subseteq P \setminus \{p\}$. Yet the proof proceeds in a similar manner. The only major difference is that the resulting similarity function may not be bounded.)

We will now show that for every H , the function $s_H(p, \cdot)$ provided by Theorem 1 is proportional to $s(p, \cdot)$ on the intersection of their domains, namely H . Let there be given any nonempty $H \subseteq P \setminus \{p\}$. (The case $H = \emptyset$ is trivial.) It will be helpful to explicitly state two lemmata:

LEMMA 1. For every $q \in H$, $s_H(p, q) = 0$ iff $s_{H^0}(p, q) = 0$.

Proof. Use axiom A5 with $H^1 = H^0$, $H^2 = H$, $q_1 = q_2 = q$, $x = y = 0$ (in \mathcal{A}^{H^0}), $z = w = 0$ (in \mathcal{A}^H) and $\alpha \neq \beta = 0$.

LEMMA 2. For every $q_1, q_2 \in H$, with $s_H(p, q_1) > 0$,

$$\frac{s_H(p, q_2)}{s_H(p, q_1)} = \frac{s_{H^0}(p, q_2)}{s_{H^0}(p, q_1)}.$$

Proof. Use axiom A5 with $H^1 = H^0$, $H^2 = H$, $x = y = 0$ (in \mathcal{A}^{H^0}), $z = w = 0$ (in \mathcal{A}^H), $\alpha = s_{H^0}(p, q_2)$ and $\beta = s_{H^0}(p, q_1)$.

We now turn to define the coefficients $\lambda_{p,H}$. Distinguish between two cases.

CASE 1. $s_H(p, \cdot) \equiv 0$.

In this case, by Lemma 1, $s(p, q) = s_{H^0}(p, q) = 0$ for every $q \in H$. Hence any $\lambda_{p,H} > 0$ will do.

CASE 2. $s_H(p, q) > 0$ for some $q \in H$.

In this case, choose such a q , and define $\lambda_{p,H} = [s_{H^0}(p, q) / s_H(p, q)] > 0$. By Lemmata 1 and 2, $\lambda_{p,H}$ is well defined.

Furthermore, it satisfies

$$\lambda_{p,H} s_H(p,q) = s_{H^0}(p,q) = s(p,q) \quad \text{for all } q \in H.$$

This completes the proof of Theorem 2.

APPENDIX 2: AN ALTERNATIVE MODEL

In this appendix we outline the axiomatic derivation of case-based decision theory where the similarity function depends not only on the problems encountered in the past, but also (potentially) on the acts chosen in these problems. We also show that V-maximization can be axiomatically derived.

We assume that the sets $P, A, R, C,$ and M are defined and interpreted as in Section II. Given a problem $p \in P$ and memory $M \subseteq C,$ we define the sets of (problem,act) pairs encountered, the set of problems encountered, and the set of acts chosen, respectively, to be

$$E = E(M) = \{(q,a) \mid \exists r \in R, (q,a,r) \in M\}$$

$$H = H(M) = \{q \in P \mid \exists a \in A, (q,a) \in E\}$$

and

$$B = B(M) = \{a \in A \mid \exists q \in P, (q,a) \in E\}.$$

For each $a \in B,$ let H_a denote the set of problems in which a was chosen; i.e., $H_a = \{q \in H \mid (q,a) \in E\}.$ Let F_a be the set of *hypothetical acts*, i.e., all the act profiles an actual act a could have had: $F_a = \{x \mid x: H_a \rightarrow \mathcal{R}\}.$ (Again, we identify the set of outcomes R with the real line and implicitly assume that it is measured in utilities.) We assume that $|B| \geq 2$ and define $F = \bigcup_{a \in B} F_a.$

For every p, E (with $|B| \geq 2$) we assume that $\geq_{p,E} \subseteq F \times F$ is a binary relation satisfying the following axioms. For simplicity of notation, $\geq_{p,E}$ will also be denoted by \geq whenever possible.

A1'. ORDER. \geq is reflexive and transitive, and for every $a, b \in B, a \neq b, x \in F_a,$ and $y \in F_b, x \geq y$ or $y \geq x.$

A2'. CONTINUITY AND COMPARABILITY. For every $a, b \in B, a \neq b$ and every $x \in F_a,$ the sets $\{y \in F_b \mid y > x\}$ and $\{y \in F_b \mid x > y\}$ are nonempty and open (in F_b endowed with the standard topology).

A2' entails a "continuity" requirement by stipulating that these sets be open; the fact that they are also assumed nonempty is

an Archimedean condition which guarantees that the similarity function will not vanish on H_a for any $a \in B$.

A3'. MONOTONICITY. For every $a, b \in B$, $a \neq b$, $x, z \in F_a$, and $y \in F_b$, if $x \geq z$ then $z \geq y$ implies that $x \geq y$, and $y \geq x$ implies that $y \geq z$.

A4'. SEPARABILITY. For every $a, b \in B$, $a \neq b$, $x, z \in F_a$, and $y, w \in F_b$, if $z \approx w$, then $x \geq y \Leftrightarrow (x + z) \geq (y + w)$.

THEOREM A2.1. The following two statements are equivalent:

- (i) for every p and E , $\geq \cdot = \geq_{p,E}$ satisfies A1'-A4';
- (ii) for every p and E there exists a function $s = s_{p,E} : H \rightarrow \mathcal{R}_+$ such that

—for all $a \in B$, $\sum_{q \in H_a} s(q) > 0$;
and
—for all $a \neq b$, $x \in F_a$, $y \in F_b$

$$x \geq_{p,E} y \Leftrightarrow \sum_{q \in H_a} s(q)x(q) \geq \sum_{q \in H_b} s(q)y(q).$$

Furthermore, in this case, for every p and E , the function $s = s_{p,E}$ is unique up to multiplication by a positive scalar.

Note that A1' requires that $\geq \cdot$ be transitive, which implies that acts belonging to the same space F_a be comparable. However, if $|B| \geq 3$, one may start out by assuming that transitivity holds only if all pairs compared belong to different spaces, and then consider the transitive closure of the original relation.

Proof (Outline). Fix $a \in B$, and consider the restriction of $\geq \cdot$ to F_a . It is easy to see that on F_a , $\geq \cdot$ is complete (hence a weak order), continuous and monotone (in the weak sense, i.e., $x \geq z$ implies that $x \geq z$).

Finally, if $x, y, z \in F_a$, we get

$$x \geq y \quad \text{iff} \quad (x + z) \geq (y + z).$$

We therefore conclude that for every $a \in B$ there is a function $s_a : H_a \rightarrow \mathcal{R}_+$ such that for all $x, y \in F_a$,

$$x \geq y \Leftrightarrow \sum_{q \in H_a} s_a(q)x(q) \geq \sum_{q \in H_a} s_a(q)y(q).$$

Furthermore, by A2', $\geq \cdot$ is nontrivial on each F_a , hence for some $q \in H_a$, $s_a(q) > 0$.

Thus, we have a numerical, additively separable representa-

tion of $\geq \cdot$ on each F_a separately. To obtain a global representation, comparing act profiles of different acts, we need to “calibrate” the various similarity functions $\{s_a\}_{a \in A}$, each of which is unique up to a positive multiplicative scalar.

One natural way to perform this calibration is to compare “constant” act profiles. That is, let I_a denote the element of F_a consisting of I 's only ($I_a(q) = 1$ for all $q \in H_a$). Fix $a \in B$, and for each $b \in B$ let δ_b satisfy

$$I_a \approx \cdot \delta_b I_b.$$

Define $s : H \rightarrow \mathcal{R}_+$ by

$$s(q) = \frac{s_b(q)}{\delta_b \sum_{q' \in H_b} s_b(q')}$$

for all $q \in H_b$. Observe that s is proportional to s_b on H_b for each $b \in B$. Thus, the s -weighted utility represents $\geq \cdot$ on F_b . Furthermore, the calibration above guarantees that, if $x \in F_a$ and $y \in F_b$ are two constant act-profiles,

$$x \geq \cdot y \Leftrightarrow \sum_{q \in H_a} s_a(q)x(q) \geq \sum_{q \in H_b} s_b(q)y(q).$$

To see that this representation holds in general, one may find for each act profile $x \in F_a$ a constant act profile $\bar{x} \in F_a$ such that $x \approx \cdot \bar{x}$ and complete the proof using transitivity of $\geq \cdot$.

Finally, it is straightforward to verify that the axioms are also necessary, and that the similarity function is unique up to a positive multiplicative scalar.

To obtain the representation by the functional V in (**), consider the following axiom.

A6. EXPERIENCE INVARIANCE. For all $a, b \in B$, $I_a \approx \cdot I_b$.

Without judging its reasonability, we note that A6 means that the “quality” of the experience is all that matters, rather than its “quantity.” Specifically, imposing A6 on top of A1'–A4' guarantees that in the construction of s above, $\delta_b = 1$ for all $b \in B$. Thus, if two acts always yielded the same result, they would be equivalent, regardless of the number of times each was chosen. Preferences satisfying A6 focus on the “average performance” of each act, disregarding any accumulated measures of performance. In other words, A1'–A4' and A6 are necessary and sufficient conditions on $\geq \cdot$ to be representable by a functional V as in (**) for given p, E . (To

obtain the V representation using a single similarity function $s(p,q)$ for all sets E , one needs to impose an additional axiom corresponding to A5.) Formally,

COROLLARY A2.2. The following two statements are equivalent:

- (i) for every p and E , $\geq \cdot = \geq_{p,E}$ satisfies A1'–A4' and A6;
- (ii) for every E there exists a function $s_E : H^c \times H \rightarrow \mathcal{R}_+$ such that for every $p \notin H$:
 - for all $a \in B$, $\sum_{q \in H_a} s_E(p,q) = 1$;
 - and
 - for all $a \neq b$, $x \in F_a$, and $y \in F_b$,

$$x \geq_{p,E} y \Leftrightarrow \sum_{q \in H_a} s_E(p,q)x(q) \geq \sum_{q \in H_b} s_E(p,q)y(q).$$

Furthermore, in this case, for every E , the function s_E is unique.

Note that the average performance as measured by V may still be a weighted average. One may further demand that $\geq \cdot$ satisfy the following axiom:

A7. CONSTANT SIMILARITY. For every $a, b \in B$, $a \neq b$, $q, q' \in H_a$, $x \in F_b$,

$$v_q \geq \cdot x \quad \text{iff} \quad v_{q'} \geq \cdot x,$$

where $v_q, v_{q'}$ stand for the corresponding unit vectors in F_a .

It is rather straightforward to show that A1'–A4', A6 and A7 are necessary and sufficient conditions for $\geq \cdot$ to be representable by a simple average. Specifically, if a preference order which is representable by V also satisfies A7, the intrinsic similarity function s in (**) (before normalization) is constant, and the functional V reduces to the simple average utility each act yielded in the past. Formally,

COROLLARY A2.3. The following two statements are equivalent:

- (i) for every p and E , $\geq \cdot = \geq_{p,E}$ satisfies A1'–A4', A6 and A7;
- (ii) for every p and E , for all $a \neq b$, $x \in F_a$, $y \in F_b$,

$$x \geq_{p,E} y \Leftrightarrow \frac{\sum_{q \in H_a} x(q)}{|H_a|} \geq \frac{\sum_{q \in H_b} y(q)}{|H_b|}.$$

REFERENCES

- Allais, M., "Le Comportement de L'Homme Rationel devant le Risque: Critique des Postulats et Axioms de l'Ecole Americaine," *Econometrica*, XXI (1953), 503-46.
- Anscombe, F. J., and R. J. Aumann, "A Definition of Subjective Probability," *The Annals of Mathematics and Statistics*, XXXIV (1963), 199-205.
- Bewley, T., "Knightian Decision Theory: Part I," mimeo, 1986.
- Camerer, C., and M. Weber, "Recent Developments in Modeling Preferences: Uncertainty and Ambiguity," *Journal of Risk and Uncertainty*, V (1992), 325-70.
- Cross, J. G., *A Theory of Adaptive Economic Behavior* (New York: Cambridge University Press, 1983).
- de Finetti, B., "La Prevision: Ses Lois Logiques, Ses Sources Subjectives," *Annales de l'Institut Henri Poincaré*, VII (1937), 1-68; translated by H. E. Kyburg in Kyburg and Smokler (1964).
- Ellsberg, D., "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics*, LXXV (1961), 643-69.
- Falkenhainer, B., K. D. Forbus, and D. Gentner, "The Structure-Mapping Engine: Algorithmic Example," *Artificial Intelligence*, XLI (1989), 1-63.
- Gick, M. L., and K. J. Holyoak, "Analogical Problem Solving," *Cognitive Psychology*, XII (1980), 306-55.
- Gick, M. L., and K. J. Holyoak, "Schema Induction and Analogical Transfer," *Cognitive Psychology*, XV (1983), 1-38.
- Gilboa, I., and D. Schmeidler, "Maxmin Expected Utility with a Non-Unique Prior," *Journal of Mathematical Economics*, XVIII (1989), 141-53.
- Gilboa, I., and D. Schmeidler, "Case-Based Consumer Theory," mimeo, 1993a.
- Gilboa, I., and D. Schmeidler, "Act Similarity in Case-Based Decision Theory," mimeo, 1994a.
- Gilboa, I., and D. Schmeidler, "Case-Based Knowledge Representation," mimeo, 1994b.
- Gilboa, I., and D. Schmeidler, "Case-Based Optimization," *Games and Economic Behavior*, forthcoming.
- Gittins, J. C., "Bandit Processes and Dynamic Allocation Indices," *Journal of the Royal Statistical Society, B*, XLI (1979), 148-64.
- Harless, D., and C. Camerer, "The Utility of Generalized Expected Utility Theories," *Econometrica*, LXII (1994), 1251-89.
- Hume, D., *Enquiry into the Human Understanding* (Oxford: Clarendon Press, 1748, second ed., 1966).
- Keynes, J. M., *A Treatise on Probability* (London: Macmillan and Co., 1921).
- Knight, F. H., *Risk, Uncertainty, and Profit* (Boston, New York: Houghton Mifflin, 1921).
- Machina, M., "Choice Under Uncertainty: Problems Solved and Unsolved," *Economic Perspectives*, I (1987), 121-54.
- March, J. G., and H. A. Simon, *Organizations* (New York: John Wiley and Sons, 1958).
- Matsui, A., "Expected Utility Theory and Case-Based Reasoning," mimeo, 1993.
- Ramsey, F. P., "Truth and Probability," in *The Foundations of Mathematics and Other Logical Essays* (London: Kegan Paul; and New York: Harcourt, Brace and Co., 1931).
- Riesbeck, C. K., and R. C. Schank, *Inside Case-Based Reasoning* (Hillsdale, NJ: Lawrence Erlbaum Associates, 1989).
- Savage, L. J., *The Foundations of Statistics* (New York: John Wiley and Sons, 1954).
- Schank, R. C., *Explanation Patterns: Understanding Mechanically and Creatively* (Hillsdale, NJ: Lawrence Erlbaum Associates, 1986).
- Schmeidler, D., "Subjective Probability and Expected Utility without Additivity," *Econometrica*, LVII (1989), 571-87.
- Selten, R., "The Chain-Store Paradox," *Theory and Decision*, IX (1978), 127-58.
- Simon, H. A., *Models of Man* (New York: John Wiley and Sons 1957).
- Tversky, A., "Features of Similarity," *Psychological Review*, LXXXIV (1977), 327-52.
- von Neumann, J., and O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton, NJ: Princeton University Press, 1944).