Economic Theory 9, 47–61 (1997)

Economic Theory

Act similarity in case-based decision theory*

Itzhak Gilboa[1] **and David Schmeidler**[2]

[1] KGSM-MEDS, Northwestern University, Leverone Hall, Evanston, IL 60208, USA

[2] Tel Aviv University and Ohio State University, Department of Economics, Arps Hall, 1945 N. High St., Columbus, OH 43210, USA

Received: May 9, 1995; revised version October 4, 1995

Summary. Case-Based Decision Theory (CBDT) postulates that decision making under uncertainty is based on analogies to past cases. In its original version, it suggests that each of the available acts is ranked according to its own performance in similar decision problems encountered in the past.

The purpose of this paper is to extend CBDT to deal with cases in which the evaluation of an act may also depend on past performance of different, but similar acts. To this end we provide a behavioral axiomatic definition of the similarity function over problem-act pairs (and not over problem pairs alone, as in the original model).

We propose a model in which preferences are context-dependent. For each conceivable history of outcomes (to be thought of as the "context" of decision) there is a preference order over acts. If these context-dependent preference relations satisfy our consistency-across-contexts axioms, there is an essentially unique similarity function that represents these preferences via the (generalized) CBDT functional.

JEL Classification Numbers: D80, D81, D83.

1. Introduction

Motivation

"Case-Based Decision Theory" (CBDT) suggests that people making decisions under uncertainty tend to choose acts that performed well in similar decision situations in the past. More specifically, the theory in its original version assumes that a decision maker has "cases" in her memory, each of which is a triple (q, a, r), where q is a decision problem, a is the act chosen in it, and r is the result that was

* We are grateful to Akihiko Matsui for the discussions that motivated this work. We also thank Enriqueta Aragones, Roger Myerson, Zvika Neeman, Ariel Rubinstein, Peyton Young, and an anonymous referee for their comments. Partial financial support from the Alfred Sloan Foundation is gratefully acknowledged.

Correspondence to: I. Gilboa

obtained. CBDT assumes a utility function over the set of results, u, and a similarity function over the set of problems, s, such that, given a memory (i.e., a set of cases) M and a decision problem p, each act a is evaluated according to the weighted sum

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

While it stands to reason that past performance of an act would affect the act's evaluation in current problems, it is not necessarily the case that past performance is the only relevant factor in the evaluation process. Specifically, an act's desirability may be affected by the performance of other, *similar* acts. For instance, suppose that a hypothetical decision maker, Ms. A, is looking for a house to buy. One of her options is to purchase a house in a neighborhood where she owned a house in the past. Ms. A hasn't lived in the same house she is now considering, yet it seems unavoidable that her past experience with a close, and probably similar house would influence her evaluation of the new one. Similarly, consider a decision maker, Mr. B, who tries to decide whether or not to buy a new product in the supermarket. He has never purchased this product in the past, but he has consumed similar products by the same producer. Again, we would expect Mr. B's decision to depend on his experience with similar acts.

More generally, a decision maker is often faced with new acts, that is, acts that her memory contains no information about their past performance. According to the CBDT formulation (Gilboa and Schmeidler, 1995), the valuation index attached to these acts is the default value zero. As in the examples above, this application of CBDT is not very plausible. Correspondingly, it may lead to counter-intuitive behavioral predictions. For instance, it would suggest that Ms. A will be as likely to buy the house in the neighborhood she knows as a house in a neighborhood she does not know. Similarly, Mr. B's decision will be predicted to buy the new product based on B's aspiration level alone, without distinguishing among products by the record of their producers.

Act similarity effects are especially pronounced in economic problems involving a continuous parameter. For instance, the decision whether or not to "Offer to sell at price p" for a specific value p, would likely be affected by the results of previous offers to sell with different but close values of p. Generally, if there are infinitely many acts available to the decision maker, it is always the case that most of them are "new" to her. However, she will typically infer something about these "new" acts from the performance of other acts she has actually tried. While a straightforward application of CBDT to economic models with an infinite set of acts may result in counter-intuitive and unrealistic predictions, the introduction of act similarity may improve these predictions. Furthermore, some of the results obtained in Gilboa and Schmeidler (1993a) and Gilboa and Schmeidler (1993c) for a finite set of acts are likely to have natural extensions to cases with infinitely many acts, provided some notion of similarity over the latter.

The need for modeling act-similarity may sometimes be obviated by redefining "acts" and "problems." For instance, Ms. A's acts may be simply "To Buy" and "Not to Buy," where each possible purchase is modeled as a separate decision problem.

However, such a model is hardly very intuitive, especially when many acts are considered simultaneously. It is more natural to explicitly model a similarity function between acts. Moreover, in many cases the similarity function is most naturally defined on problem-act pairs. For example, "Driving on the left in New York" may be similar to "Driving on the right in London;" "Buying when the price is low," may be more similar to "Selling when the price is high" than to "Selling when the price is low," and so forth.

In short, we would like to have a model in which the similarity function s is defined on such pairs, and – again, given a memory M and a decision problem p, – each act a is evaluated according to the weighted sum

$$(\bullet) \qquad U(a) = U_{p,M}(a) = \sum_{(q,b,r)\in M} s((p,a),(q,b))u(r).$$

Observe that a case (q, b, r) in memory may be viewed as a pair $((q, b), r)$, where the problem-act pair (q, b) is a single entity, describing the circumstances from which the outcome r resulted. That is, when past cases are considered, the distinction between problems and acts is immaterial. Indeed, it may also be fuzzy in the decision maker's memory. By contrast, when evaluating currently available acts, this distinction is both clearer and more important: the "problem" refers to the given circumstances, which are not under the decision maker's control, whereas the various "acts" describe alternative choices.

Axiomatization

In this paper we provide an axiomatization of the decision rule ($\bullet$). The axiomatization is different in nature from those typically found in the literature, as well as from the axiomatization of ($*$) we present in Gilboa and Schmeidler (1995). In the latter, an act is identified with its own past history. That is, each act can be thought of as a function, associating an outcome with each problem in which this act was chosen. Thus the mathematical structure we use there resembles that of de Finetti (1937) and, to an extent, also that of Savage (1954) and Anscombe and Aumann (1963). Namely, the objects of choice are vectors of outcomes. True, in the case-based model any two distinct acts are defined on disjoint domains, whereas in the other models all acts have the same domain. Yet the spirit of the axioms in Gilboa and Schmeidler (1995) is similar to the classical models of expected utility theory.

By contrast, when one takes into account the effect of act similarity, one can no longer identify an act with its own history of outcomes. The latter does not summarize all the information relevant to the evaluation of the act. Rather, the outcome of any act in any problem may, a-priori, affect the evaluation of any other act in the current decision problem.

We therefore use a model in which preferences are context-dependent. For each conceivable history of outcomes (to be thought of as the "context" of decision) there is a preference order over acts. Our axioms relate these preference relations given different contexts. A decision maker whose preferences across contexts is "consistent" enough to satisfy our axioms can be ascribed an essentially-unique problem-act similarity function such that her preferences are represented by ($\bullet$).

The axiomatization of (•) may be viewed as a recipe for indirect elicitation of similarity by a direct elicitation of preferences between acts in hypothetical contexts. The decision maker facing a problem with a set of available (feasible) acts A_p has one 'true' memory $M = ((q_i, b_i, r_i))_{i=1}^n$. Our implicit assumption is that her considerations in selecting an act from A_p follow (•). That is, in evaluating an act, say a, she assigns weights to consequences in her memory, the $(u(r_i))_{i=1}^n$, by their similarities to the act being evaluated, $(s((p, a), (q_i, b_i)))_{i=1}^n$, correspondingly. When faced with a hypothetical memory $M' = ((q_i, b_i, r_i'))_{i=1}^n$ she ranks the alternatives using, we assume, *the same* similarities. Indeed, when using a hypothetical memory to elicit preferences, only the results are changed and the problems and acts of the "true" memory are left unchanged.

There is a restrictive but simplifying assumption in our axiomatization. The results are expressed in "utiles," and the utility function is not derived simultaneously with the similarity function. In Section 3 the possibility to generalize the present work in this and other directions in presented, together with additional remarks. In the next section we describe our formal model, axioms and results. The proofs are relegated to Section 4.

Before concluding this section, it should be mentioned that Matsui (1993) also provides an axiomatization of CBDT with problem-act similarity. His model allows the similarity function to depend on past cases' outcomes as well. The axiomatic derivation of Matsui presupposes a binary relation on functions that assign an outcome to triples of the form (past case, act, act). From a mathematical viewpoint, his model and results resemble those of Gilboa and Schmeidler (1995), but the objects of choice he uses are naturally more complex entities than the "act profiles" used in the latter. There appears to be no simple relationship between Matsui's results and those of the present paper. While both papers axiomatize similar functionals, they employ rather different frameworks.

2. The axiomatization and the theorem

We start with the model's primitives, following Gilboa and Schmeidler (1995).

Let P be a non-empty set of *problems*. Let A be a non-empty set of *acts*. For each $p \in P$ there is a non-empty subset $A_p \subseteq A$ of acts available at p. Let R be a set of *outcomes* or *results*. The set of *cases* is $C \equiv P \times A \times R$. A *memory* is a finite subset $M \subseteq C$. Without loss of generality we assume that for every memory M, if $m = (p, a, r), m' = (p', a', r') \in M$ and $m \neq m'$, then $p \neq p'$. As mentioned in the introduction, we assume (explicitly) that $R = \Re$ and, implicitly, that the utility function is the identity.

To state our axioms the following notation will prove useful: given a memory M, denote its projection on $P \times A$ by E. That is,

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

is the set of problem-act pairs recalled. We will also use the projection of M (or of E) on P, denoted by H. That is,

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

is the set of problems recalled.

For every memory M and every problem $p \notin H$ we assume a preference relation over acts, $\geq_{p,M} \subseteq A_p \times A_p$. Our main result derives the numerical representation for a *given set* E and a given *problem* $p \notin H$. Let us therefore assume that E and p are given. Every memory M with $E(M) = E$ may be identified with the results it associates with the problem-act pairs, i.e., with a function $x = x(M) \in \mathfrak{R}^E$. An element $x \in \mathfrak{R}^E$ specifies the history of results, or the *context* of the decision problem p. Denoting $n = |E|$, we abuse notation and identify $\mathfrak{R}^E$ with $\mathfrak{R}^n$. Thus a relation $\geq_{p,M}$ over A_p may be thought of as a relation $\geq_x$. Moreover, we will assume that $\geq_x$ is defined for every $x \in \mathfrak{R}^n$. We define $>_x$ and $\approx_x$ to be the asymmetric and symmetric parts of $\geq_x$, as usual.

We will use the following axioms:

A1 Order: For every $x \in \mathfrak{R}^n$, $\geq_x$ is complete and transitive on A_p.

A2 Continuity: For every $\{x_k\}_{k \geq 1} \subseteq \mathfrak{R}^n$ and $x \in \mathfrak{R}^n$, and every $a, b \in A_p$, if $x_k \to x$ (in the standard topology on $\mathfrak{R}^n$) and $a \geq_{x_k} b$ for all $k \geq 1$, then $a \geq_x b$.

A3 Additivity: For every $x, y \in \mathfrak{R}^n$ and every $a, b \in A_p$, if $a >_x b$ and $a \geq_y b$, then $a >_{x+y} b$.

A4 Neutrality: For every $a, b \in A_p$, $a \approx_0 b$, $(0 = (0, \ldots, 0) \in \mathfrak{R}^n)$.

Axiom 1 is probably the most standard of all. It simply requires that, given any conceivable context, the decision maker's preference relation over acts is a weak order. Axioms 2–4 are new in the sense that they are formulated in terms of contexts, rather than in terms of acts. However, at least Axioms 2 and 3 cannot fail to remind the reader of standard axioms in the classical approach: A2 requires that preferences be continuous *in the space of contexts*. A3 states that preferences be additive in this space. That is, if both contexts x and y suggest that a is preferred to b, and at least one of the preferences is strict, then also the "sum" context $x + y$ suggests that a is strictly preferred to b. The logic of this axiom is that a context may be thought of as "evidence" in favor of one act or another. Thus, if both x and y "lend support" to choosing a over b, then so should the "accumulated evidence" $x + y$. For instance, assume that $E = \{(q_1, c), (q_2, d)\}$ and that $x = (1, 1)$ and $y = (0, -1)$. Assume that $a >_x b$. Presumably, this is so because (p, a) is, on average, more similar to (q_1, c) and to (q_2, d) than is (p, b). Further assume that $a >_y b$. This is probably due to the fact that (p, b) is more similar to (q_2, d) than (p, a) is, hence act d's failure colors the decision maker's impression of b more negatively than that of a. It seems natural that for the context $x + y = (1, 0)$ it will be true that $a >_{x+y} b$. Indeed, the act c, which (in the corresponding problem) was more similar to a succeeded, and that which was more similar to b resulted in a neutral outcome. Thus a is preferred to b.

Needless to say, A3 is one of the main axioms, and it carries most of the responsibility for the additive functional we end up with. Naturally, it cannot be any more plausible than the additive functional itself, and there are reasonable examples in which additivity fails. (See Gilboa and Schmeidler 1993b.) However, the main role of the axiomatization here is to relate the theoretical construct, "problem-act similarity," to observable preferences. Hence we do not attempt to present A3 as a "canon of rationality." While we believe it is a sensible requirement in some cases, we concede it may fail in others.

The utility level 0 in Axiom 4 represents a neutral result. Intuitively, it makes the decision maker neither sad nor happy. We refer to this utility as the "aspiration level." If all problem-act pairs in the decision maker's memory resulted in this neutral utility level, she cannot use her memory to differentiate between available acts. Any act, whether close to or remote from previous acts, cannot be expected to perform better than any other act. (Note that if every act in memory resulted in the same positive utility, then one would prefer an act that is, on average, closer to past acts, to a more remote act. In this case strict preferences between some acts are expected.)

Axioms 1–4 are easily seen to be necessary for the functional form we would like to derive. By contrast, the next axiom we introduce is not. While the theorem we present is an equivalence theorem, it characterizes a more restricted class of preferences than the decision rule discussed in the introduction, namely those preferences satisfying Axiom 5 as well. This axiom should be viewed merely as a technical requirement. It states that preferences are "diverse" in the following sense: for any four acts, there is a conceivable context that would distinguish among all four of them.

A5 Diversity: For every distinct $a, b, c, d \in A_p$ there exists $x \in \Re^n$ such that $a >_x b >_x c >_x d$.

(Observe that A5 is trivially satisfied when $|A_p| < 4$.)

Note that, specifically, A5 rules out preferences according to which acts c and d are always "between" a and b. This may be particularly restrictive for some applications. For instance, consider acts that are linearly ordered, say, they are parametrized by quantity. In this case it may well be the case that "Sell 100 shares" is preferred to "Sell 300 shares," or vice versa – but that in both cases, "Sell 200 shares" is ranked in between the two. Yet (in the presence of at least four acts), this is precluded by A5. Therefore there is certainly room to study more general axiom systems. In the next section we discuss Axiom 5 in more detail and provide examples to show that Axioms 1–4 alone cannot guarantee the desired result.

Our main result can now be formulated.

Theorem: Let there be given E and p as above. Then the following two statements are equivalent:

(i) $\{\geq_x\}_{x \in \Re^n}$ satisfy A1–A5;

(ii) There are vectors $\{s^a\}_{a \in A_p}$, $s^a \in \Re^n$, such that:
for every $x \in \Re^n$ and every $a, b \in A_p$,

(**)
$$a \geq_x b \qquad \text{iff} \qquad \sum_{i=1}^{n} s_i^a x_i \geq \sum_{i=1}^{n} s_i^b x_i,$$

and, for every distinct $a, b, c, d \in A_p$, the vectors $(s^a - s^b)$, $(s^b - s^c)$ and $(s^c - s^d)$ are linearly independent.

Furthermore, in this case, if $|A_p| \geq 4$, the vectors $\{s\}_{a \in A_p}$ are unique in the following sense: if $\{s^a\}_{a \in A_p}$ and $\{\hat{s}^a\}_{a \in A_p}$ both satisfy (**), then there are a scalar $\alpha > 0$ and a vector $w \in \Re^n$ such that for all $a \in A_p$, $\hat{s}^a = \alpha s^a + w$.

We remind the reader that $\Re^n$ is used as a proxy for $\Re^E$. Thus the vectors $\{s^a\}_{a \in A_p}$ provided by the theorem can also be thought of as functions from E to $\Re$. Furthermore, these can be viewed as defining similarity on problem-act pairs. Specifically, the theorem implies that under A1–A5, there exists a similarity function

$$s_E: \{(p,a) | p \notin H(E), a \in A_p\} \times E \to \Re$$

defined by

$$s_E((p,a),(q,b)) = s^a((q,b)),$$

such that for every p, M with $E(M) = E$ and $p \notin H(M)$ the functional

$$U_{p,M}(a) = \sum_{(q,b,r) \in M} s_E((p,a),(q,b))u(r)$$

represents $\geq_{p,M}$ on A_p.

Since the formulation in (•) did not include explicitly the dependence of the similarity function on the pairs of problem-act recalled, E, the following question naturally arises: What additional condition (or "axiom") we need to impose so that the similarity between pairs (p,a) and $(q,b) \in E$ would be independent of E? The answer is deferred to the next section.

3. Remarks

Remark 1 (Aspiration level adjustment): The decision rule (•) is generally not invariant under shifts of the utility function. The utility level zero has been interpreted as the "aspiration level," and different aspiration levels would lead to different choices. Moreover, a decision maker's aspiration level need not be constant over time. If we adopt a cognitive interpretation of this behaviorally-defined concept, it is plausible that the payoff people "count on" getting in a given problem depends on the payoff they experienced in the past. Correspondingly, the way in which aspiration levels are adjusted may have important behavioral implications. For instance, in Gilboa and Schmeidler (1993c) we show that an aspiration level adjustment rule that satisfies certain conditions will asymptotically lead to optimal choices in a repeated problem. More generally, there is a need for a theory of aspiration level adjustments.

We do not offer here any such theory. Yet our main theorem axiomatizes a class of decision rules that includes both (•), and variations thereof allowing the aspiration level to change over time, as long as it is a linear function of the payoffs received in the past. Specifically, consider a functional

$$U_{p,M}(a) = \sum_{(q,b,r) \in M} s_E((p,a),(q,b))\,[u(r) - h(x)]$$

where $h(x)$ is linear, i.e., $h(x) = \sum_{i=1}^{n} \alpha_i x_i$ for some $\alpha_i \in \Re$, $(i = 1, \ldots, n)$, and x is the vector of payoff, as above. It is easily verified that in this case one can re-define the similarity function so that the above expression reduces to that axiomatized in the theorem. In other words, as long as we allow the similarity function to depend on E, the functional form we axiomatize is general enough to describe at least a certain class of aspiration level adjustment rules.

Remark 2 (Insufficiency of A1–4): As mentioned above, only A1–4 are necessary for the numerical representation we axiomatize. One may wonder whether we can make do without A5 altogether. We note here that this is not the case, i.e., that axioms 1–4 are not sufficient for the numerical representation (•). We prove this by two counter-examples in the next section. The first is combinatorial in nature, and uses only four acts. The second is based on cardinality of the set of acts. Specifically, note that axioms A1–4 do not restrict the act set in terms of cardinality, topology, and so forth. Hence, as this example shows, A1–4 are compatible with a lexicographic ordering of a large set of acts, and therefore cannot suffice for a numerical representation.

Remark 3 (Memory independence): The condition guaranteeing that the similarity function is independent of memory will be expressed here in terms of the similarity function provided by the representation theorem. Alternatively, one may use the language of the original data, namely the preference orders $\geq_{p,M} \subseteq A_p \times A_p$. However, it may make the following condition even more cumbersome, with no obvious theoretical advantage.[1]

For every $p, q, q' \in P$, every $a, b, c, a', b', c' \in A_p$ and every E^1 and E^2 such that (q, b), $(q', b') \in E^1, E^2$, and $p \notin H(E^1), H(E^2)$, the following holds:

$$\frac{s_{E^1}((p,a'),(q',b')) - s_{E^1}((p,c'),(q',b'))}{s_{E^1}((p,a),(q,b)) - s_{E^1}((p,c),(q,b))} = \frac{s_{E^2}((p,a'),(q',b')) - s_{E^2}((p,c'),(q',b'))}{s_{E^2}((p,a),(q,b)) - s_{E^2}((p,c),(q,b))}$$

whenever the denominators do not vanish. It is further assumed that if one denominator equals zero, so does the other. If the condition holds, then $s_{E^1} = s_{E^2}$.

The proof follows that of the corresponding result – Theorem 2 – in Gilboa and Schmeidler (1995).

Remark 4: For some purposes, one might be interested in a model with an infinite memory, and a functional that is a similarity-weighted integral over it. Since all our arguments are based on duality (or separation) results, the axiomatic derivation of such a similarity measure, together with the corresponding functional, is relatively straightforward.

Remark 5: From a theoretical point of view, the concept of "utility," as that of "similarity," should also be related to observable choices. Specifically, one would like to have an axiomatic derivation of both concepts simultaneously. One way to obtain such a derivation would be to first derive a notion of comparison of utility differences from choices among acts, and then to require that our additivity axiom hold with respect to these differences. Using a cardinal utility function that

[1] Since the similarity function is almost uniquely determined by the main theorem, it is already translated to observable data. Note that the condition that follows does not depend on the specific choice of the similarity function provided by the theorem.

represents the difference comparisons as in Alt (1936), we may then proceed with our proof in the utility space. Although we do not provide such an axiomatization here, we point out a possible way to infer utility differences comparisons from act preference data.

Note that in the decision rule (•), the pair of similarity and utility functions (s, u) is equivalent to the pair $(-s, -u)$. Intuitively, observing a preference of act a over b given a history of act c, it is possible that c was a success, and a (in the current problem) is more similar to it (in the recalled problem) than b is, but also that c was a failure, and a is *less* similar to it than is b.

One may therefore assume a given preference order over the set of outcomes R. Using it, a qualitative similarity order on *pairs* of problem-act pairs is defined. One may then proceed to define a binary relation on pairs of results, to be interpreted as strength of preferences. Thus, $(r, r') \geq (t, t')$ if for all $a, b \in A_p$, and all M_1, M_2, M_3, M_4 with $E(M_i) = E, i = 1, \ldots, 4$: If M_1 and M_2 as well as M_3 and M_4 (viewed as functions from E to R) differ only on (q, c), $(\bar{q}, \bar{c}) \in E$; (q, c, t), $(\bar{q}, \bar{c}, s) \in M_1$, (q, c, t'), $(\bar{q}, \bar{c}, s') \in M_2$, (q, c, r), $(\bar{q}, \bar{c}, s) \in M_3$ and (q, c, r'), $(\bar{q}, \bar{c}, s') \in M_4$; (q, c) is more similar to (p, a) than to (p, b), and the outcome r is preferred to r'; and $a \approx_{M_1} b$, $a \approx_{M_2} b$ and $a \approx_{M_3} b$, then $a \leq_{M_4} b$. One then imposes a few axioms on this binary relation on $R \times R$ that guarantee a cardinal utility representation.

Remark 6: The axiomatization we propose here assumes that preference relations are given for memories that differ from the actual one in the results obtained (but not in the problem-act pairs encountered). It was pointed out to us by Peyton Young and by Roger Myerson that one may also provide an axiomatization of our decision rule assuming that preferences are given for memories that are derived from the actual one by replication of cases. That is, such memories would have a different number of cases than the actual one, but each of these cases will be identical to a real case. The formal structure of such a derivation follows closely the derivations of scoring rules by Young (1975) and by Myerson (1993).

Specifically, for a given (actual) memory M, let the set of conceivable memories be N^M, where N stands for the non-negative integers. Summation of elements in N^M, as well as multiplication by an integer, are interpreted pointwise. Let $\phi: \mathrm{N}^M \rightarrow \rightarrow A$ be a non-empty-valued correspondence. We interpret ϕ as the observable choice correspondence, selecting the "best" acts in A for every conceivable memory. We seek axioms on ϕ that will be equivalent to the existence of "weights" $w(c, a)$ for every case $c \in M$ and every act $a \in A$, such that, for every conceivable memory $\Theta \in \mathrm{N}^M$,

(◇)
$$\phi(\Theta) = \arg\max_{a \in A} \left[\sum_{c \in M} \Theta(c) w(c, a) \right].$$

That is, $w(c, a)$ is the weight assigned to case c in the evaluation of act a in the present problem. In our model, $w(c, a) = s((p, a), (q, b))u(r)$ where $c = (q, b, r)$ and p is the problem at hand. However, (◇) need not be interpreted as a product of a similarity and a utility function.

The Young-Myerson axioms would be:

Additivity: For all Θ, $\Xi \in \mathrm{N}^M$ with $\phi(\Theta) \cap \phi(\Xi) \neq \varnothing$, $\phi(\Theta + \Xi) \subseteq \phi(\Theta) \cap \phi(\Xi)$.

(Archimedean) Continuity: For all Θ, $\Xi \in \mathrm{N}^M$ there is a K such that for all $n \geq K$, $\phi(n\Theta + \Xi) \subseteq \phi(\Theta)$.

Diversity: For every non-empty $B \subseteq A$, there is a $\Theta \in \mathrm{N}^M$ such that $\phi(\Theta) = B$.

This model has the conceptual advantage that all conceivable memories, for which preferences are assumed to be given, are obtained from the actual one by replicating actual cases. On the other hand, for the very same reason such a model cannot differentiate between the effects of past problems, of the acts chosen at them, and of the results that followed. In particular, one cannot obtain a separate axiomatic derivation of the similarity function and of the utility function; rather, only their product is observable.

4. Proofs

We split the proof into three parts: (i) implies (ii), the opposite direction, and the uniqueness of the representation in part (ii). Finally, we prove Remark 2.

Part 1: *(i) implies (ii)*

To outline the proof we will first state its main 3 lemmata.

Lemma 1: There are vectors $\{s^{ab}\}_{a,b \in A_p}$, $s^{ab} \in \Re^n$, such that

(i) $X^{ab} \equiv \{x \in \Re^n \mid a \geq_x b\} = \{x \in \Re^n \mid s^{ab} \cdot x \geq 0\}$;
(ii) $Y^{ab} \equiv \{x \in \Re^n \mid a >_x b\} = \{x \in \Re^n \mid s^{ab} \cdot x > 0\}$;
In particular, $Y^{ab} = \varnothing$ iff $s^{ab} = 0$;
(iii) $-s^{ab} = s^{ba}$;
(iv) s^{ab} satisfying (i), and (ii), is unique up to multiplication by a positive number.

The proof uses Axioms 1–4 to show that the sets X^{ab} and Y^{ab} are convex and a separation theorem can therefore be applied.

Lemma 2: For every three acts, $f, g, h \in A_p$, and the corresponding vectors s^{fg}, s^{gh}, s^{fh} from Lemma 1 there are α, $\beta \geq 0$ such that:

(i) $\alpha s^{fg} + \beta s^{gh} = s^{fh}$.
(ii) Moreover, if the acts are distinct, and $|A_p| \geq 4$, then α, β from (i) are unique.

The proof of (i) uses the linear programming duality result. In the proof of (ii) Axiom 5 is used for the first time; it also is required for the proof of the following lemma.

Lemma 3: Suppose that $|A_p| \geq 4$, then there are vectors $\{s^{ab}\}_{a,b \in A_p}$, $s^{ab} \in \Re^n$ such that: (i)–(iii) of Lemma 1 hold, and for any three acts, f, g, $h \in A_p$, the Jacobi identity $s^{fg} + s^{gh} = s^{fh}$ holds.

Note that Lemma 3, unlike Lemma 2, guarantees the possibility to rescale *simultaneously all* the s^{ab}-s from Lemma 1 such that the Jacobi identity will hold on A_p.

Lemma 3 is proved by double induction; it is transfinite if A_p is uncountably infinite. Before proving the lemmata, we will show how Part (i) is derived from Lemma 3.

Proof of Part (i): First let us point out, omitting details, that for the case $|A_p| \leq 2$ Lemma 1 implies Part (i), and for the case $|A_p| = 3$, Lemma 2(i) implies Part (i). Hence we restrict attention to the case $|A_p| \geq 4$.

Choose an arbitrary act, say e in A_p. Define $s^e = 0$, and for any other act, a, define $s^a = s^{ae}$, where the s^{ae}-s are from Lemma 3.

Given $x \in \Re^n$ and $a, b \in A_p$ we have:

$$a \geq_x b \Leftrightarrow s^{ab} \cdot x \geq 0 \Leftrightarrow$$
$$(s^{ae} + s^{eb}) \cdot x \geq 0 \Leftrightarrow (s^{ae} - s^{be}) \cdot x \geq 0 \Leftrightarrow s^a \cdot x - s^b \cdot x \geq 0 \Leftrightarrow s^a \cdot x \geq s^b \cdot x$$

The first implication follows from Lemma 1(i), the second from the Jacobi identity of Lemma 3, the third from Lemma 1(iii), and the fourth from the definition of the s^a-s. Hence, (**) of the theorem has been proved.

Given four distinct acts $a, b, c, d \in A_p$, suppose by way of negation, that the vectors $(s^a - s^b)$, $(s^b - s^c)$ and $(s^c - s^d)$ are linearly dependent. I.e.,

$$\alpha(s^a - s^b) + \beta(s^b - s^c) + \gamma(s^c - s^d) = 0,$$

where one coefficient is positive and at least another one is nonnegative. As an example set $\alpha < 0$, $\beta > 0$, $\gamma \geq 0$. Then,

$$s^a - s^b = \left(-\frac{\beta}{\alpha}\right)(s^b - s^c) + \left(-\frac{\gamma}{\alpha}\right)(s^c - s^d).$$

Here, applying (**), for any $x \in \Re^n$: if $b >_x c$ and $c >_x d$, then $a >_x b$. This precludes the ranking $b >_x c >_x d >_x a$ (for any x), the existence of which is guaranteed by A5. Such a contradiction can be found for any of the remaining possible combinations of signs of the coefficients α, β and γ. ◇

Proof of Lemma 1: We first state and prove four claims that are immediate implications of the axioms.

Claim 1, Reflection: $a \geq_x b \Rightarrow b \geq_{-x} a$

Proof: Otherwise, by A1, $a >_{-x} b$. This, together with the antecedent of the claim and A3 contradict A4. ◇

Claim 2, Reflection (continued): $a >_x b \Rightarrow b >_{-x} a$

Proof: Otherwise Claim 1 is contradicted. ◇

Claim 3 Additivity (continued): $a \geq_x b$ and $a \geq_y b$ imply $a \geq_{x+y} b$.

Proof: Otherwise, by A1 and Claim 2, $a >_{-(x+y)} b$. Since $-(x+y)+y=x$, we get by A3; $a >_{-x} b$. In view of Claim 1, this contradicts $a \geq_x b$. ◇

Claim 4 Homogeneity: For $\lambda > 0$, $a \geq_x b$ iff $a \geq_{\lambda x} b$.

Proof: Prove first for λ an integer, then for λ a rational number and complete by A2.

Completion of the Proof of Lemma 1: Applying Axioms 1–4 and Claims 1–4 we get that X^{ab} and Y^{ab} are convex cones with vertex at the origin. Also X^{ab} is closed, Y^{ab} is open and $X^{ba} = -X^{ab}$, $Y^{ba} = -Y^{ab}$.

We will now construct the s^{ab}-s for all unordered pairs and singletons $\{\{a,b\}|a,b\in A_p\}$. If Y^{ba}, the complement of X^{ab}, is empty, then define $s^{ab}=0=s^{ba}$. Clearly, (i), (ii), (iii), hold for this case. If Y^{ba} is non-empty then the origin is a boundary point of X^{ab}. By a supporting hyperplane argument there is a vector $s^{ab}\neq 0$ so that (i) holds. Define $s^{ba}=-s^{ab}$; thus, (iii) holds. Obviously (ii) holds too. By Claims 1 and 2 (Reflection), conditions (i) and (ii) also hold for (b,a). (iv) is implied by (i). ◇

Proof of Lemma 2:

(i) Given three vectors s^{fg}, s^{gh}, s^{fh} define the following LP problem:

$$(P)\qquad \begin{aligned} &\mathrm{Min}_{x\in\Re^n} s^{fh}\cdot x\\ &s.t.\quad s^{fg}\cdot x\geq 0\\ &\qquad\quad s^{gh}\cdot x\geq 0.\end{aligned}$$

Consider also its dual problem (with a degenerate objective function),

$$(D)\qquad \begin{aligned} &\mathrm{Max}_{\alpha,\beta} 0\alpha+0\beta\ \ (\equiv 0)\\ &s.t.\quad \alpha s^{fg}+\beta s^{gh}=s^{fh}\\ &\qquad\quad \alpha,\beta\geq 0.\end{aligned}$$

Transitivity of $\geq_x$ implies that (P) is bounded: every x that is feasible for (P) satisfies $f\geq_x g$ and $g\geq_x h$. Hence it also satisfies $f\geq_x h$, which implies that $s^{fh}\cdot x\geq 0$. Since (P) is bounded, (D) is feasible. It follows that there are α, $\beta\geq 0$ such that $\alpha s^{fg}+\beta s^{gh}=s^{fh}$. ◇

(ii) Assume by way of negation, that for some distinct f, g, $h\in A_p$: $\alpha s^{fg}+\beta s^{gh}=s^{fh}=\gamma s^{fg}+\delta s^{gh}$, or $(\alpha-\gamma)s^{fg}=(\delta-\beta)s^{gh}$, where $(\alpha,\beta)\neq(\gamma,\delta)$. By Lemma 1, this implies that either $X^{fg}=X^{gh}$ or $X^{fg}=-X^{gh}$. The first case precludes the existence of x for which $f>_x h>_x g$ and the second case precludes the existence of x such that $f>_x g>_x h$. In both cases A5 is violated. ◇

Proof of Lemma 3: Assume that A_p is (well-) ordered. Let acts i,j and k be the first three acts in this order. Let s^{ik} be a vector from Lemma 1, and apply Lemma 2 to the three acts, i,j,k. Define $\hat{s}^{ij}=\alpha s^{ij}$ and $\hat{s}^{jk}=\beta s^{jk}$. (Recall that α and β are unique, given s^{ik}). Supplement these and later definitions by $s^{aa}=0$, $\forall a\in A_p$, and by condition (iii) of Lemma 1. Clearly, the conclusions of Lemma 3 hold for all $f,g,h\in\{i,j,k\}$. For simplicity of presentation we rename the vectors, $\hat{s}^{ij}$ and $\hat{s}^{jk}$ by s^{ij} and s^{jk}, respectively.

Continue by induction where the induction hypothesis is stated as follows: Suppose that for some two acts, say c and d, with c preceding d, the vectors s^{ab} satisfying the conclusion of Lemma 3 have been defined for all pairs of acts a,b preceding d as well as for pairs of acts a, d and d, a with a preceding c. Suppose also that s^{dc} has not yet been defined.

If there is no such a d, then the proof of the Lemma has been completed. If there is no such a c, then the main induction step has been completed and we move to the act immediately following d in the order. If $c=i$, then we start the secondary induction step: To define s^{di} and s^{dj} we apply Lemma 2, as above, to i, d, j, using s^{ij} as given. Obviously the conclusions of the Lemma hold on the extended domain.

We present now the induction's main step. To define s^{dc} apply Lemma 2 to the three acts i, d, c with s^{ic} from the induction hypothesis. We get a uniquely defined s^{dc}, but we get also a new definition of s^{id}, to be denoted by λs^{id}. I.e., the Jacobi identity from Lemma 2 in the present notation is:

$$\lambda s^{id} + s^{dc} = s^{ic}$$

On the other hand, for any act preceding c, say a, we have by the induction hypothesis,

$$s^{id} + s^{da} = s^{ia}$$

Applying once again Lemma 2 to the three acts a, d, c we get,

$$\vartheta s^{ad} + \eta s^{dc} = s^{ac}.$$

Subtracting the last two equalities from the previous one we get,

$$(\lambda - 1)s^{id} + (1 - \eta)s^{dc} + (-\vartheta - 1)s^{ad} = s^{ic} - s^{ia} - s^{ac} = 0$$

The right side equality follows from the Jacobi identity for the three acts i, a, c by the induction hypothesis.

As a conclusion we have that either the vectors s^{id}, s^{dc} and s^{ad} are linearly dependent or all three coefficients are zero. In the latter case it has been shown that $\lambda = 1$ and adding the vector s^{ad} (and $s^{da} = -s^{ad}$ and $s^{dd} = 0$) preserves the Jacobi identity and other conclusions of Lemma 3 on the extended domain. I.e., the induction step has been proved in this case, and we can pass to the immediate follower of c (in the order), if it differs from d.

In the linear dependence case we get a contradiction to Axiom 5. Indeed, it implies that a vector in one of the pairs $\{\{s^{id}, s^{di}\}, \{s^{dc}, s^{cd}\}, \{s^{ad}, s^{da}\}\}$ can be represented as nonnegative linear combination of two vectors, one from each of the remaining pairs. Any such representation leads to a contradiction, as in the Proof of Part (i) above. For example, if s^{dc} is a nonnegative linear combination of s^{di} and s^{ad}, then for no x: $c >_x a >_x d >_x i$. ◇

Part 2: (ii) implies (i)

It is straightforward to verify that if $\{\geq_x\}_{x \in \mathfrak{R}^n}$ are representable by $\{s^a\}_{a \in A_p}$ as in (**), they have to satisfy Axioms 1–4. We will therefore only prove that this representation – coupled with the linear independence condition – imply Axiom 5.

Assume, then, that some quadruple of distinct acts $a, b, c, d \in A_p$ is given, and that the vectors $(s^a - s^b)$, $(s^b - s^c)$ and $(s^c - s^d)$ are linearly independent. We will prove that there exists an $x \in \mathfrak{R}^n$ such that $a >_x b >_x c >_x d$. Because of the linear independence, the following linear system has a solution:

$$(s^a - s^b) \cdot x = 1$$
$$(s^b - s^c) \cdot x = 1$$
$$(s^c - s^d) \cdot x = 1.$$

The desired order is implied by (**). ◇

Part 3: *Uniqueness*

Suppose that $|A_p| \geq 4$ and that $\{s^a\}_{a\in A_p}$ and $\{\hat{s}^a\}_{a\in A_p}$ both satisfy (**), and we wish to show that there are a scalar $\alpha > 0$ and a vector $w \in \Re^n$ such that for all $a \in A_p$, $\hat{s}^a = \alpha s^a + w$. Recall that, for $a \neq b$, $s^a \neq s^b$ and $\hat{s}^a \neq \hat{s}^b$ by A5.

Choose $a \neq b$ $(a, b \in A_p)$. From Lemma 1(iv) there exists a unique $\alpha > 0$ such that $(\hat{s}^a - \hat{s}^b) = \alpha(s^a - s^b)$. Define $w = \hat{s}^a - \alpha s^a$.

We now wish to show that, given $c \in A_p$: $\hat{s}^c = \alpha s^c + w$. Again, from Lemma 1(iv) there are unique γ, $\delta > 0$ such that

$$(\hat{s}^c - \hat{s}^a) = \gamma(s^c - s^a)$$

and

$$(\hat{s}^b - \hat{s}^c) = \delta(s^b - s^c).$$

Summing up these two with $(\hat{s}^a - \hat{s}^b) = \alpha(s^a - s^b)$, we get

$$\alpha(s^a - s^b) + \gamma(s^c - s^a) + \delta(s^b - s^c) = 0.$$

Thus

$$(\alpha - \delta)(s^a - s^b) + (\delta - \gamma)(s^a - s^c) = 0.$$

Since $(s^a - s^b)$ and $(s^b - s^c)$ are linearly independent (condition (ii) of the Theorem), so are $(s^a - s^b)$ and $(s^a - s^c)$, and we conclude that $\alpha = \gamma = \delta$. This implies $\hat{s}^c = \alpha(s^c - s^a) + \hat{s}^a = \alpha s^c + w$. ◇

This completes the proof of the theorem.

Proof of Remark 2: We will now show that A1–4 are not sufficient for (•) to hold. As mentioned above, we provide two counter-examples. Each highlights a different aspect of A5.

Example 1: Let $A_p = \{a, b, c, d\}$ and $n = 3$. Define the vectors

$$s^{ab} = (-1, 1, 0) \qquad s^{ac} = (0, -1, 1); \qquad s^{ad} = (1, 0, -1);$$
$$s^{bc} = (2, -3, 1); \qquad s^{cd} = (1, 2, -3); \qquad s^{bd} = (3, -1, -2),$$

and $s^{uv} = -s^{vu}$ and $s^{uu} = 0$ for $u, v \in A_p$.

For $u, v \in A_p$ and $x \in \Re^3$ define, $u \geq_x v$ iff $s^{uv} \cdot x \geq 0$.

We claim that $\{\geq_x\}_{x\in\Re^3}$ satisfy A1–4 and yet they cannot be represented as in (•). With the exception of transitivity, A1–4 follow immediately from the definition of $\{\geq_x\}_{x\in\Re^3}$, and Lemma 1. Transitivity is verified triple by triple. We now show that the numerical representation (•) does not hold.

Assume by way of negation that there are vectors $\{s^u\}_{u\in A_p}$ representing $\{\geq_x\}_{x\in\Re^3}$. Define $\hat{s}^{uv} = s^u - s^v$. By Lemma 2, for every $u, v \in A_p$ there exists a coefficient $\lambda^{uv} > 0$ such that $\hat{s}^{uv} = \lambda^{uv} s^{uv}$. Without loss of generality we may assume $\lambda^{bc} = 1$. Since s^{bc} and s^{cd} are linearly independent, we also get $\lambda^{cd} = \lambda^{db} = 1$.

Considering the equation $\lambda^{ab} s^{ab} + \lambda^{ca} s^{ca} = s^{cb}$, observe that it has a unique solution with $\lambda^{ab} = 2$. (Where uniqueness follows from linear independence of s^{ab} and s^{ca}.) Hence we have $\hat{s}^{ab} = (-2, 2, 0)$. By a similar token, the equation $\lambda^{da} s^{da} + \lambda^{ab} s^{ab} = s^{db}$ also has a unique solution in which $\lambda^{ab} = 1$. Thus we also have $\hat{s}^{ab} = (-1, 1, 0)$, a contradiction.

Example 2: Let $A_p = [0, 1]^2$ and let $\geq_l$ be the lexicographic order on it. For any given $n \geq 1$, define $\{\geq_x\}_{x \in \mathfrak{R}^n}$ as follows:

$$\text{if } \sum_{i=1}^{n} x_i > 0, \quad a >_x b \qquad \text{iff} \quad a >_l b;$$

$$\text{if } \sum_{i=1}^{n} x_i = 0, \quad a \approx_x b \qquad \text{for all } a, b \in A_p;$$

and

$$\text{if } \sum_{i=1}^{n} x_i < 0, \quad b >_x a \qquad \text{iff} \quad a >_l b.$$

Thus, for every $x \in \mathfrak{R}^n$, $\geq_x$ is one of (i) $\geq_l$; (ii) $\geq_l^{-1}$; or (iii) the trivial relation (according to which any two acts are equivalent). Hence $\leq_x$ satisfies A1. It can also be verified that A2–A4 are satisfied by $\{\geq_x\}_{x \in \mathfrak{R}^n}$. However, one would not expect to obtain a representation as in (∗∗), since it would imply a numerical representation of $\geq_l$ as well.

We therefore conclude that A5, which is also used in our proof for the finite case, implicitly bounds the cardinality of the set of acts A_p. Specifically, $|A_p| \leq \aleph$ since there cannot be more than a continuum of independent vectors in $\mathfrak{R}^n$. ◇

References

Alt, F.: On the measurability of utility. In: John S. Chipman, Leonid Hurwicz, Marcel K. Richter, Hugo F. Sonnenschein (eds.) Preferences, Utility, and demand, A. Minnesota Symposium, pp. 424–431, chapt. 20. New York, Chicago, San Francisco, Atlanta: Harcourt Brace Jovanovich, Inc., [Über die Messbarheit des Nutzens. Zeitschrift fuer Nationaloekonomie **7**, 161–169 (1936), Abstract (in English): On the Measurability of Utility. Translation by Siegfried Schach, revised and corrected by Dr. Alt]

de Finetti, B.: La prevision: ses lois logiques, ses sources subjectives. Annales de l'Institute Henri Poincare **7**, 1–68 (1937)

Gilboa, I., Schmeidler, D.: Case-based decision theory. The Quarterly Journal of Economics *110*, 605–639 (1995)

Gilboa, I., Schmeidler, D.: Case-based consumer theory. Mimeo., Northwestern University, 1993a, revised 1994

Gilboa, I., Schmeidler, D.: Case-based knowledge representation. Mimeo, Northwestern University, 1993b, revised 1994

Gilboa, I., Schmeidler, D.: Case-based optimization. Games and Economic Behavior, (forthcoming)

March, J. G., Simon, H. A.: Organizations. New York: Wiley 1958

Matsui, A.: Expected utility theory and case-based reasoning. Mimeo, University of Pennsylvania, 1993

Myerson, R. B.: Axiomatic derivation of scoring rules without the ordering assumption. Social Choice and Welfare (forthcoming)

Ramsey, F. P.: Truth and probability. In: The foundations of mathematics and other logical essays. New York: Harcourt Brace Jovanovich, Inc. 1931. Reprinted in Kyburg and Smokler 1964

Savage, L. J.: The foundations of statistics, 2nd edn. New York: John Wiley and Sons 1972

Simon, A.: Models of man. New York: Wiley 1957

Young, P. H.: Social choice scoring function. SIAM Journal of Applied Mathematics **28**, 824–838 (1975)