A Festschrift for Herman Rubin
Institute of Mathematical Statistics
Lecture Notes – Monograph Series
Vol. 45 (2004) 45–55

A Rubinesque theory of decision

J. B. Kadane[1], Mark J. Schervish[1], and Teddy Seidenfeld[1]

Carnegie Mellon University

Abstract: We generalize a set of axioms introduced by Rubin (1987) to the case of partial preference. That is, we consider cases in which not all uncertain acts are comparable to each other. We demonstrate some relations between these axioms and a decision theory based on sets of probability/utility pairs. We illustrate by example how comparisons solely between pairs of acts is not sufficient to distinguish between decision makers who base their choices on distinct sets of probability/utility pairs.

1. Introduction

Rubin (1987) presented axioms for rational choice amongst sets of available actions. These axioms generalize those of Von Neumann and Morgenstern (1947) which deal solely with comparisons between pairs of actions. Both of these sets of axioms imply that all actions that are choosable from a given set are equivalent in the sense that the rational agent would be indifferent between choosing amongst them. We weaken the axioms of Rubin (1987) by allowing that the agent might not be able to choose between actions without being indifferent between them.

There are several reasons for allowing noncomparbility (unwillingness to choose without being indifferent) between actions. One simple motivation is a consideration of robustness of decisions to changes in parts of a statistical model. For example, consider an estimation problem with a loss function but several competing models for data and/or parameters. We might be interested in determining which estimators can be rejected in the sense that they do not minimize the expected loss under even a single one of the competing models. The agent may not be indifferent between the estimators that remain without being able to select a best one.

With regard to sets of choices, (Rubin, 1987, p. 49) says "The basic concept is that of a *choice set*. This is a set of actions that will be chosen by decision maker; we do *not* assume the decision maker can select a unique action." Nevertheless, the axioms of Rubin (1987) lead to a unique (up to positive affine transformation) utility that ranks all actions, just as do the axioms of Von Neumann and Morgenstern (1947). The weakening of the axioms that we present here is consistent with a set of utilities combined through a Pareto-style criterion, which we introduce in Section 3.

2. Comparison of axioms

Initially, we consider a nonempty convex collection $\mathcal{A}$ of *acts*. In particular, for every $x_1, x_2 \in \mathcal{A}$ and every $0 < a < 1$, $ax_1 + (1-a)x_2 \in \mathcal{A}$. As such, the set of acts must lie in some part of a space where convex combination makes sense. Typically, we think of acts either as probability distributions over a set $\mathcal{R}$ or as functions from some other set Ω to probability distributions on $\mathcal{R}$. These interpretations make convex combination a very natural operation, but the various axiom systems and the related theorems do not rely on one particular class of interpretations.

[1]Department of Statistics, Carnegie Mellon University, Pittsburgh, Pittsburgh, PA 15213, USA. e-mail: kadane@stat.cmu.edu; mark@stat.cmu.edu; teddy@stat.cmu.edu
Keywords and phrases: partial preference.
AMS 2000 subject classifications: primary 62C05; secondary 62A01.

The classic axioms of Von Neumann and Morgenstern (1947) are the following.

Von Neumann–Morgenstern Axiom 1. There exists a weak order $\preceq$ on $\mathcal{A}$. That is,

- for every $x \in \mathcal{A}$, $x \preceq x$,
- for every $x, y \in \mathcal{A}$, either $x \preceq y$, or $y \preceq x$, or both, and
- for all $x, y, z \in \mathcal{A}$, if $x \preceq y$ and $y \preceq z$, then $x \preceq z$.

In the case in which $x \preceq y$ and $y \preceq x$, then we say $x \sim y$.

Von Neumann–Morgenstern Axiom 2. For all $x, y, z \in \mathcal{A}$, $x \preceq y$ if and only if for all $0 < a \leq 1$ $ax + (1-a)z \preceq ay + (1-a)z$.

Von Neumann–Morgenstern Axiom 2 is the most controversial of the classic axioms. Its appeal stems from the following scenario. Imagine that a coin (independent of everything else in the problem) is flipped with probability a of landing heads. If the coin lands heads, you must choose between x and y, otherwise, you get z. Presumably the choice you would make between x and y would be the same in this setting as it would be if you merely had to choose between x and y without any coin flip. The controversy arises out of the following scenario. The coin flip that determines which of x or z arises from $ax+(1-a)z$ can be different (although with the same probability) from the coin flip that determines which of y or z arises from $ay+(1-a)z$. From a minimax standpoint, the first scenario can lead to a different choice between $ax+(1-a)z$ and $ay+(1-a)z$ than does the second scenario.

Von Neumann–Morgenstern Axiom 3. For all $x, y, z \in \mathcal{A}$, if $x \preceq y \preceq z$, then there exists $0 \leq a \leq 1$ such that $y \sim ax + (1-a)z$.

Von Neumann–Morgenstern Axiom 3 prevents any acts from being worth infinitely more (or infinitesimally less) than other acts. Under these axioms, Von Neumann and Morgenstern (1947) prove that there exists a utility $U : \mathcal{A} \to \mathbb{R}$ satisfying

- for all $x, y \in \mathcal{A}$, $x \preceq y$ if and only if $U(x) \leq U(y)$,
- for all $x, y \in \mathcal{A}$ and $0 < a < 1$, $U(ax + (1-a)y) = aU(x) + (1-a)U(y)$, and
- U is unique up to positive affine transformation.

The axioms of Rubin (1987), which we state next, make use of the convex hull of a set $E \subseteq \mathcal{A}$ which is denoted $H(E)$. Rubin (1987) was particularly concerned with the idea that, when presented with a set E of actions, the agent might insist on randomizing between actions in E rather than selecting an action from E itself. This is why the choice set from E is a subset of $H(E)$.

Rubin Axiom 1. There is a function $C : 2^{\mathcal{A}} \to 2^{\mathcal{A}}$ that satisfies

- for all $E \in 2^{\mathcal{A}}$ $C(E) \subseteq H(E)$, and if E has 1, 2, or 3 elements then $C(E) \neq \emptyset$.

The set C in Rubin Axiom 1 can be thought of as a generalization of the weak order $\preceq$ from Von Neumann–Morgenstern Axiom 1: $x \preceq y$ if and only if $y \in C(\{x, y\})$.

Rubin Axiom 2. For all $T, S \in 2^{\mathcal{A}}$, if $T \subseteq H(S)$ and $H(T) \cap C(S) \neq \emptyset$, then $C(T) = H(T) \cap C(S)$.

Rubin Axiom 2 says that if an act is choosable from a large set, then it remains choosable from any smaller set that contains it.

If $S \subseteq \mathcal{A}$, $x \in \mathcal{A}$, and $0 \leq a \leq 1$, define $aS + (1-a)x = \{ay + (1-a)x : y \in S\}$.

Rubin Axiom 3. For all $S \subseteq \mathcal{A}$ and all $0 < a < 1$, if $C(S) \neq \emptyset$, then $C(aS + (1-a)x) = aC(S) + (1-a)x$.

Rubin Axiom 3 is the obvious analog to Von Neumann–Morgenstern Axiom 2.

Rubin Axiom 4. Let $S \subseteq \mathcal{A}$ and $x \in H(S)$. If, for all $V \subseteq H(S)$, ($x \in V$ and $C(V) \neq \emptyset$) implies $x \in C(V)$, then $x \in C(S)$.

Rubin Axiom 4 says that, if an act is not choosable from S, then it is not choosable from some subset of S.

Rubin Axiom 5. Let $x, y, z \in \mathcal{A}$ be such that $C(\{x, y\}) = \{x\}$ and $C(\{y, z\}) = \{y\}$. Then there exists $0 < a < 1$ such that $\{y, ax+(1-a)z\} \subseteq C(\{y, ax+(1-a)z\})$.

Rubin Axiom 5 is an obvious analog to Von Neumann–Morgenstern Axiom 3. Under these axioms, Rubin (1987) proves that there exists a utility $U : \mathcal{A} \to \mathbb{R}$ satisfying

- for all $E \subseteq \mathcal{A}$, $C(E) = \{x \in H(E) : \text{ for all } y \in E, U(x) \geq U(y)\}$,
- for all $x, y \in \mathcal{A}$ and $0 < a < 1$, $U(ax + (1-a)y) = aU(x) + (1-a)U(y)$, and
- U is unique up to positive affine transformation.

It is fairly simple to show that, if such a U exists, then all of Rubin's axioms hold. Hence, his result is that his axioms characterize choice sets that are related to utility functions in the way described by the three bullets above.

In order to allow noncomparability, we need more general axioms than Von Neumann–Morgenstern Axiom 3 and Rubin Axiom 5. To state the more general axioms, we need a topology on the set of actions. For now, assume that the set of acts $\mathcal{A}$ is a metric space with some metric d. When we consider specific examples, we will construct the metric. Let $\mathcal{F}$ be the collection of nonempty closed subsets of $\mathcal{A}$.

We prefer to state our axioms in terms of a rejection function rather than a choice set function.

Definition 1. A *rejection set* R is a function $R : \mathcal{F} \to 2^{\mathcal{A}}$ such that, for all $E \in \mathcal{F}$, $R(E) \subseteq E$, and $R(E) \neq E$.

Axiom 1. If $B \subseteq R(A)$ and if $A \subseteq D$, then $B \subseteq R(D)$.

Axiom 1 is the same as Sen's property α. (see Sen (1977)). It says that adding more options to a set of acts doesn't make the rejected ones become acceptable.

Axiom 2. If B is a subset of $R(A)$ and if D is a subset of B, then $B \backslash D \subseteq R(\overline{A \setminus D})$.

Axiom 2 says that rejected acts remain rejected even if we remove other rejected acts from the option set.

Definition 2. For $A, B \in \mathcal{F}$, say that $A \prec B$ if $A \subseteq R(A \cup B)$.

Lemma 1. *Assume Axiom 1 and Axiom 2. Then $\prec$ is a strict partial order on $\mathcal{F}$.*

Proof. Let $A, B \in \mathcal{F}$. If $A \prec B$, then $B \not\prec A$ because A and B being closed implies that $R(A \cup B) \neq A \cup B$. For transitivity, assume that $A \prec B$ and $B \prec D$ with $A, B, D \in \mathcal{F}$. Then $A \subseteq R(A \cup B) \subseteq R(A \cup B \cup D)$, by Axiom 1. Also Axiom 1 says that $B \subseteq R(B \cup D) \subseteq R(A \cup B \cup D)$. It follows that $A \cup B \subseteq R(A \cup B \cup D)$. Let $E = B \setminus A$. Then

$$A = (A \cup B) \setminus E \subseteq R(\overline{[A \cup B \cup D] \setminus E}) \subseteq R(A \cup D),$$

where the first inclusion is from Axiom 2 and the second is from Axiom 1. □

Our next axiom is similar to Rubin Axiom 3.

Axiom 3. For all $E \in \mathcal{F}$, all $x \in \mathcal{A}$, and all $0 < a \leq 1$, $B = R(E)$ if and only if $aB + (1-a)x = R(aE + (1-a)x)$.

For the continuity axiom, we require the concept of a sequence of sets that are all indexed the same way.

Definition 3. Let G be an index set with cardinality less than that of $\mathcal{A}$. Let H be another index set. Let $\mathcal{H} = \{E_h : h \in H\}$ be a collection of subsets of $\mathcal{A}$. We say that the sets in $\mathcal{H}$ are *indexed in common by* G if For each $h \in H$ and each $g \in G$, there exists $x_{h,g} \in \mathcal{A}$ such that $E_h = \{x_{h,g} : g \in G\}$.

Axiom 4. Let G_A and G_B be index sets with cardinalities less than that of $\mathcal{A}$. Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of elements of $\mathcal{F}$ such that each $A_n = \{x_{n,g} : g \in G_A\}$. Also, let $\{B_n\}_{n=1}^{\infty}$ be a sequence of elements of $\mathcal{F}$ such that each $B_n = \{x_{n,g} : g \in G_B\}$. Suppose that for each $g \in G_A \cup G_B$, $x_{n,g} \Rightarrow x_g \in \mathcal{A}$. Let $A = \{x_g : g \in G_A\}$ and $B = \{x_g : g \in G_B\}$. Let N and J be closed subsets of $\mathcal{A}$.

- If $\forall n B_n \prec A_n$ and $\overline{A} \prec N$, then $\overline{B} \prec N$.
- If $\forall n B_n \prec A_n$ and $J \prec \overline{B}$, then $J \prec \overline{A}$.

The reason for wording Axiom 4 with the additional sets N and J is that the acts in $\overline{B}$ and $\overline{A}$ might be noncomparable when compared to each other because the limit process brings B_n and A_n so close together. But the axiom says that the limit of a sequence of rejected options can't jump over something that is better than the limit of choosable options. Similarly, the limit of a sequence of choosable options can't jump below something that is worse than the limit of rejected options.

We state one additional axiom here that is necessary for the generalization that we hope to achieve. Recall that $H(E)$ is the convex hull of the set E.

Axiom 5. For each $E \in \mathcal{F}$ and $B \subseteq E$, if $B \subseteq R(\overline{H(E)})$, then $B \subseteq R(E)$.

Axiom 5 says that if acts are rejected when the closed convex hull of E is available then they must also be rejected when E alone is available. Closing the convex hull of a closed set of acts should not allow us to reject any acts that we couldn't reject before.

3. Pareto Criteria

After proving the existence of the utility, Rubin (1987) considers cases with many utilities indexed by elements of some set Ω. He then says (p. 53) "Two immediate examples come to mind: Ω may be the class of states of nature, or Ω may be the set of all individuals in a population. Suppose we assume that the choice process given ω is 'reasonable' for each $\omega \in \Omega$, and the overall process is also reasonable." The first of the two examples envisioned by Rubin (1987) is the usual case in which there is uncertainty about unknown events. The second example is the case in which the "overall process" is governed by a social welfare function. Our approach is motivated by an alternative way of thinking about individuals in a population. Instead of a social welfare function that performs just like an individual's utility, we seek a characterization of the agreements amongst the individuals.

Definition 4. Let $\aleph$ be a set. For each $\alpha \in \aleph$, let $R_\alpha : \mathcal{F} \to 2^{\mathcal{A}}$ be a rejection function. The *Pareto rejection function related to* $\{R_\alpha : \alpha \in \aleph\}$ is $R(E) = \bigcap_{\alpha\in\aleph} R_\alpha(E)$ for all $E \in \mathcal{F}$.

In this definition, an act x is *Pareto rejected* by the group $\aleph$ if it is rejected by every member of the group. The complement of the Pareto rejection function might be called the Pareto choice function $C : \mathcal{F} \to 2^{\mathcal{A}}$ defined by $C(E) = [R(E)]^C$. This is the set of acts that fail to be rejected by at least one individual in $\aleph$.

The general example that motivates our work is the following. Let Ω be a finite set of states. For each $\alpha \in \aleph$ let P_α be a probability on Ω. Let acts be functions from Ω to probability measures over some finite set of prizes $\mathcal{R}$. That is, let $\mathcal{P}_{\mathcal{R}}$ be the set of probability measures over $\mathcal{R}$ so that each act $x \in \mathcal{A}$ is a function $x : \Omega \to \mathcal{P}_{\mathcal{R}}$ and $x(\omega)(r)$ is the probability of prize r in state ω. For each $\alpha \in \aleph$, let there be a bounded possibly state-dependent utility $U_\alpha(\cdot|\omega)$, appropriately measurable. Define $V_\alpha : \mathcal{A} \to \mathbb{R}$ by

$$V_\alpha(x) = \sum_{\omega\in\Omega} \left[\sum_{r\in\mathcal{R}} U_\alpha(r|\omega) x(\omega)(r) \right] P_\alpha(\omega).$$

Next, define

$$C_\alpha(E) = \big\{ x \in E : V_\alpha(x) \geq V_\alpha(y), \text{ for all } y \in E \big\},$$

and $R_\alpha(E) = [C_\alpha(E)]^C$. Hence, C_α is the set of all Bayes rules in the model with utility U_α and probability P_α. Then $R(E) = \bigcap_{\alpha\in\aleph} R_\alpha(E)$ is the set of all acts x such that, for every model in $\aleph$, x fails to be a Bayes rule. We call this rejection function the *Bayes rejection function related to* $\{(P_\alpha, U_\alpha) : \alpha \in \aleph\}$. Finally, we define the metric on $\mathcal{A}$. All acts are equivalent to points in a bounded subset of a finite-dimensional Euclidean space. If s is the number of states and t is the number of prizes, then act x is equivalent to an $s \times t$ matrix with (i, j) entry equal to the probability of prize j in state i. We will use the usual Euclidean metric as d. It is now easy to see that V_α is a continuous function of x for each α.

Lemma 2. *If $B \prec A$, then, for each $\alpha \in \aleph$, there is $y \in A \setminus B$ such that $V_\alpha(y) > \sup_{z\in B} V_\alpha(z)$.*

Proof. We can think of B as a closed and bounded subset of a finite-dimensional Euclidean space. For each $\alpha \in \aleph$, V_α is continuous, hence there exists $x \in B$ such that $V_\alpha(x) = \sup_{z\in B} V_\alpha(z)$. Since $x \in B$, there exists $y \in A \cup B$ such that $V_\alpha(x) < V(y)$. By the definition of x it is clear that $y \in A \setminus B$. ☐

Lemma 3. *The Bayes rejection function related to $\{(P_\alpha, U_\alpha) : \alpha \in \aleph\}$ satisfies Axiom 1.*

Proof. Let $A \in \mathcal{F}$ and $B \subseteq R(A)$ and $A \subseteq D$. If $x \in B$, then for each $\alpha \in \aleph$, there is $y_\alpha \in A$ such that $V_\alpha(x) < V_\alpha(y_\alpha)$. Since $y_\alpha \in D$ for all α, it follows that $x \in R(D)$. ☐

Lemma 4. *The Bayes rejection function related to* $\{(P_\alpha, U_\alpha) : \alpha \in \aleph\}$ *satisfies Axiom* 2.

Proof. Let B be a closed subset of $R(A)$ and let $D \subseteq B$. Let $x \in B \setminus D$. Since $x \in B$, for every $\alpha \in \aleph$, there exists $y_\alpha \in A \setminus B$ such that $V_\alpha(x) < V_\alpha(y_\alpha)$ by Lemma 2. Since $y_\alpha \in A \setminus D$ as well, we have $x \in R(\overline{A \setminus D})$. $\square$

Lemma 5. *The Bayes rejection function related to* $\{(P_\alpha, U_\alpha) : \alpha \in \aleph\}$ *satisfies Axiom* 3.

Proof. The "if" direction is trivial because $a = 1$ is included. For the "only if" direction, let $0 < a \leq 1$, $x \in \mathcal{A}$ and $E \in \mathcal{F}$. First, we show that $R(aE + (1-a)x) \subseteq aR(E) + (1-a)x$. Let $z \in R(aE + (1-a)x)$. Express $z = ay + (1-a)x$, with $y \in E$. For every $\alpha \in \aleph$ there is $z_\alpha = ay_\alpha + (1-a)x$ with $y_\alpha \in E$ and $V_\alpha(z_\alpha) > V_\alpha(z)$. This implies $V_\alpha(y_\alpha) > V_\alpha(y)$ and $y \in R(E)$, so $z \in aR(E) + (1-a)x$. Finally, let $z \in aR(E) + (1-a)x$, and express $z = ay + (1-a)x$, with $y \in R(E)$. For every $\alpha \in \aleph$, there is $y_\alpha \in E$ such that $V_\alpha(y_\alpha) > V_\alpha(y)$ so that $V_\alpha(ay + (1-a)x) > V_\alpha(z)$. It follows that $z \in R(aE + (1-a)x)$. $\square$

Lemma 6. *The Bayes rejection function related to* $\{(P_\alpha, U_\alpha) : \alpha \in \aleph\}$ *satisfies Axiom* 4.

Proof. Assume $B_n \prec A_n$ for all n. Let $g \in G_B$ and $\alpha \in \aleph$. For each n, there is $h_{n,g} \in G_A$ such that $V_\alpha(x_{n,g}) < V_\alpha(x_{n,h_{n,g}})$. By continuity of V_α, we have

$$V_\alpha(x_g) \leq \liminf_n V_\alpha(x_{n,h_{n,g}}) \leq \sup_{h \in G_A} V_\alpha(x_h) \leq \sup_{x \in \overline{A}} V_\alpha(x).$$

Because V_α is continuous and $\overline{A}$ is a closed and bounded subset of a finite-dimensional Euclidean space, there exists $y \in \overline{A}$ such that $V_\alpha(y) = \sup_{x \in \overline{A}} V_\alpha(x)$. It follows that

$$\sup_{g \in G_B} V_\alpha(x_g) \leq V_\alpha(y). \tag{1}$$

For the first line of Axiom 4, assume that $\overline{A} \prec N$. For each $g \in G_B$ and each $\alpha \in \aleph$, we need to find $z \in \overline{B} \cup N$ such that $V_\alpha(x_g) < V_\alpha(z)$. Let y be as in (1). Because $\overline{A} \prec N$, there is $z \in N \setminus \overline{A} \subseteq \overline{B} \cup N$ such that $V_\alpha(z) > V_\alpha(y) \geq V_\alpha(x_g)$.

For the second line of Axiom 4, assume that $J \prec \overline{B}$. For each $x \in J$ and each $\alpha \in \aleph$, we need to find $y \in J \cup \overline{A}$ such that $V_\alpha(x) < V_\alpha(y)$. Let $x \in J$ and $\alpha \in \aleph$. By Lemma 2 there is $x_g \in \overline{B} \setminus J$ such that $V_\alpha(x) < V_\alpha(x_g)$. Let y be as in (1). Since $y \in \overline{A} \subseteq J \cup \overline{A}$, we are done. $\square$

Lemma 7. *The Bayes rejection function related to* $\{(P_\alpha, U_\alpha) : \alpha \in \aleph\}$ *satisfies Axiom* 5.

Proof. Let $E \in \mathcal{F}$ and $B \subseteq E$. Assume that $B \subseteq R(\overline{H(E)})$. Let $x \in B$. For each $\alpha \in \aleph$, we know that there exists $z_\alpha \in \overline{H(E)}$ such that $V_\alpha(x) < V_\alpha(z_\alpha)$. This z_α is a limit of elements of $H(E)$ and V_α is continuous, hence there is a $w_\alpha \in H(E)$ such that $V_\alpha(x) < V_\alpha(w_\alpha)$. This w_α is a convex combination of elements of E, $w_\alpha = \sum_{i=1}^{\ell} a_i w_{i,\alpha}$ with $w_{i,\alpha} \in E$ and $\sum_{i=1}^{\ell} a_i = 1$ with all $a_i \geq 0$. Since

$$V_\alpha(w_\alpha) = \sum_{i=1}^{\ell} a_i V_\alpha(w_{i,\alpha}) > V_\alpha(x),$$

there must exist i such that $V_\alpha(w_{i,\alpha}) > V_\alpha(x)$. Let $y_\alpha = w_{i,\alpha}$. $\square$

What the preceding results establish is that the Bayes rejection function related to a collection of probability/utility pairs satisfies our axioms. We would like to consider the opposite implication, that is, whether or not every rejection function that satisfies our axioms is the Bayes rejection function related to some collection of probability utility pairs. This consideration will be postponed until another paper.

4. Pairwise choice is not enough

Seidenfeld, Schervish and Kadane (1995) consider the first four axioms that we have introduced in this paper but restricted to the collection of subsets of the form $\{x, y\}$ with $x, y \in \mathcal{A}$. That is, Seidenfeld, Schervish and Kadane (1995) consider choices between pairs of acts only. They go on to prove that, under these axioms, there exists a collection of bounded utilities $\{V_\alpha : \alpha \in \aleph\}$ that agree with all pairwise choices in the following sense: $\{x\} \prec \{y\}$ if and only if $V_\alpha(x) < V_\alpha(y)$ for all $\alpha \in \aleph$. The following example illustrates why Axiom 5 is necessary in the case of choice between more than two acts at a time.

Example 1. Let $\mathcal{A} = \{(a, b) : 0 \leq a, b \leq 1\}$. Define the rejection function R as follows. For $(a, b) \in E$, $(a, b) \in R(E)$ if and only if there exists $(c, d) \in E$ such that, for every $0 \leq p \leq 1$, $ap + b(1-p) < cp + d(1-p)$. It is not difficult to show that this rejection function satisfies our first four axioms. However, there is no set of utility functions for which this rejection function is the Pareto rejection function. Suppose that U were an element of such a set of utility functions. By Axiom 3, $U(a, b)$ would have to equal $aU(1, 0) + bU(0, 1)$, hence

$$U(0.4, 0.4) = 0.4[U(1, 0) + U(0, 1)] < \max\{U(1, 0), U(0, 1)\}.$$

Hence either $U(1, 0) > U(0.4, 0.4)$ or $U(0, 1) > U(0.4, 0.4)$. Now, let $E = \{(0.4, 0.4), (1, 0), (0, 1)\}$, and notice that $R(E) = \emptyset$. But every utility function U would reject $(0.4, 0.4)$ amongst the actions in E.

The rejection function in Example 1 is an example of "Maximality" that was introduced by Walley (1990). The distinction between pairwise choice and larger choice sets goes beyond the situation of Example 1. Schervish, Seidenfeld, Kadane and Levi (2003) look more carefully at the special case of Bayes rejection functions in which all U_α are the same function U and $\{P_\alpha : \alpha \in \aleph\} = \mathcal{P}$, is a convex set of probabilities on Ω. We call this the case of a *cooperative team*. In this case, they give an example that illustrates how different sets $\mathcal{P}$ lead to the same collections of pairwise choices that satisfy the axioms of Seidenfeld, Schervish and Kadane (1995). Hence, pairwise choices are not sufficient for characterizing the corresponding set of probability/utility pairs even in the cases in which such sets of probability/utility pairs are known to exist.

Example 2. Let $\Omega = \{\omega_1, \omega_2, \omega_3\}$ consist of three states. Let

$$\begin{aligned}
\mathcal{P}_1 &= \big\{(p_1, p_2, p_3) : p_2 < 2p_1 \quad \text{for } p_1 \leq 0.2\big\} \\
&\quad \bigcup \big\{(p_1, p_2, p_3) : p_2 \leq 2p_1 \quad \text{for } 0.2 < p_1 \leq 1/3\big\}, \\
\mathcal{P}_2 &= \big\{(p_1, p_2, p_3) : p_2 < 2p_1 \quad \text{for } p_1 < 0.2\big\} \\
&\quad \bigcup \big\{(p_1, p_2, p_3) : p_2 \leq 2p_1 \quad \text{for } 0.2 \leq p_1 \leq 1/3\big\}.
\end{aligned}$$

The only difference between the two sets is that $(0.2, 0.4, 0.4) \in \mathcal{P}_2 \setminus \mathcal{P}_1$. Let $R_{\mathcal{P}_1}$ and $R_{\mathcal{P}_2}$ be the Bayes rejection functions corresponding to the two sets of probability/utility pairs $\{(p, U) : p \in \mathcal{P}_1\}$ and $\{(p, U) : p \in \mathcal{P}_2\}$. Each act x can be represented by the vector whose ith coordinate is $x_i = \sum_{r \in R} U(r|\omega_i)x(\omega_i)(r)$ for $i = 1, 2, 3$. In this way, the expected utility for each probability vector p is $V_p(x) = x^\top p$. Consider two arbitrary acts x and y. We have $\{x\} \in R_{\mathcal{P}_j}(\{x, y\})$ if and only if

$$\sum_{i=1}^{3}(y_i - x_i)p_i > 0, \qquad \text{for all } p \in \mathcal{P}_j.$$

This is equivalent to $(y_1 - x_1, y_2 - x_2, y_3 - x_3)$ being a hyperplane that separates $\{0\}$ from $\mathcal{P}_j$ without intersecting $\mathcal{P}_j$. It is easy to check that a hyperplane separates $\mathcal{P}_1$ from $\{0\}$ without intersecting $\mathcal{P}_1$ if and only if it separates $\mathcal{P}_2$ from $\{0\}$ without intersecting $\mathcal{P}_2$. The reason is that all of the points in the symmetric difference $\mathcal{P}_1 \Delta \mathcal{P}_2$ are extreme but not exposed. Hence, all pairwise comparisons derived from $R_{\mathcal{P}_1}$ are identical to those derived from $R_{\mathcal{P}_2}$.

Consider now a set of acts E that contains only the following three acts (each expressed as a vector of its expected payoffs in the three states as were x and y above):

$$\begin{aligned} f_1 &= (0.2, 0.2, 0.2), \\ f_2 &= (1, 0, 0), \\ g &= (-1.8, 1.2, .2). \end{aligned}$$

First, let $p \in \mathcal{P}_1$. Notice that $V_p(f_2)$ is the highest of the three whenever $p_1 \geq 0.2$, $V_p(f_1)$ is the highest whenever $p_1 \leq 0.2$, and $V_p(g)$ is never the highest. So, $R_{\mathcal{P}_1}(E) = \{g\}$. Next, notice that if $p = (0.2, 0.4, 0.4)$, then $V_p(g) = V_p(f_1) = V_p(f_2) = 0.2$, so $R_{\mathcal{P}_2}(E) = \emptyset$. $\quad\diamondsuit$

Next, we present a theorem which states that the more general framework of rejection functions operating on sets of size larger than 2 can distinguish between different sets $\mathcal{P}$ in the cooperative team case.

For the general case, let U be a single, possibly state-dependent, utility function. For each probability vector p on Ω and each act x, let

$$V_p(x) = \sum_{\omega \in \Omega}\left[\sum_{r \in R} U(r|\omega)x(\omega)(r)\right]p(\omega).$$

Because the inner sum $w_x(\omega) = \sum_{r \in R} U(r|\omega)x(\omega)(r)$ does not depend on p, we can represent each act x by the vector

$$\big(w_x(\omega_1), \ldots, w_x(\omega_s)\big), \tag{2}$$

where s is the number of states. That is, each act might as well be the vector in (2) giving for each state the state-dependent expected utility with respect to its probability distribution over prizes in that state. If we call the vector in (2) by the name x, this makes $V_p(x) = x^\top p$ for every act x and every probability vector p.

For each convex set $\mathcal{P}$ of probability vectors there is a Bayes rejection function defined by

$$R_{\mathcal{P}}(E) = \bigcap_{p \in \mathcal{P}} \big\{x \in E : V_p(x) \geq V_p(y), \text{ for all } y \in E\big\}^C, \tag{3}$$

for all closed sets E of acts. Example 2 shows that there are cases in which $\mathcal{P}_1 \neq \mathcal{P}_2$ but $R_{\mathcal{P}_1}(E) = R_{\mathcal{P}_2}(E)$ for every E that contains exactly two distinct acts. Theorem 1 below states that, so long as $\mathcal{P}_1 \neq \mathcal{P}_2$ there exists a finite set E of acts such that $R_{\mathcal{P}_1}(E) \neq R_{\mathcal{P}_2}(E)$.

Theorem 1. *Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be distinct convex sets of probabilities over a set Ω with $s \geq 2$ states. Then there is a set E with with at most $s+1$ acts such that $R_{\mathcal{P}_1}(E) \neq R_{\mathcal{P}_2}(E)$.*

The proof of Theorem 1 is given in the appendix.

5. Summary

In this paper, we consider a generalization of Subjective Expected Utility theory in which not all options are comparable by a binary preference relation. We adapt Rubin's (1987) axioms for rational choice functions to permit a decision maker who has a determinate cardinal utility U for outcomes to have a choice function over simple horse-lottery options that does not coincide with a weak ordering of the option space. In calling the decision maker's choice function "rational", we mean that there is a cardinal utility U and a set $\mathcal{P}$ of coherent probabilities that represent the choice function in the following sense: The allowed choices from an option set are exactly those Bayes-admissible options, i.e. those options that maximize expected utility for some probability $P \in \mathcal{P}$.

In Sections 2 and 3 we give axioms that are necessary for a choice function to be rational in this sense. We show that the axioms that we used in Schervish, Seidenfeld, Kadane and Levi (1995) for a theory of coherent strict partial orders are insufficient for this purpose. Specifically, those axioms are for a strict partial order $\prec$ which is given by pairwise comparisons solely. That theory represents the strict partial order $\prec$ by a set of probability/utility pairs according to a Pareto condition, where each probability/utility pair agrees with the strict partial order according expected utility inequalities. Here we show that the choice function that Walley calls "Maximality" obeys those axioms, but fails to have the desired representation in terms of Bayes-admissible options when the option sets (which may fail to be convex) involve three or more options. Therefore, we add a new Axiom 5 that is necessary for a choice function to be rational, and which is not satisfied by Maximality.

In Section 4 we show that, even when a rational choice function is represented by a convex set of coherent probabilities, and when the option set also is convex, nonetheless the choice function cannot always be reduced to pairwise comparisons. We show how to distinguish the choice functions based on any two different convex sets of probabilities using choice problems that go beyond pairwise comparisons.

In continuing work, we seek a set of axioms that characterize all rational choice functions. The axioms that we offer in Section 2 are currently a candidate for that theory.

A. Proof of Theorem 1

First, we present a few lemmas about convex sets that will be useful for the proof.

The following result gives us a way of reexpressing a half-space of a hyperplane as the intersection of the hyperplane with a half-space of a more convenient form. The main point is that the same constant c that defines the original hyperplane H can also be used to define the new half-space.

Lemma 8. *Let $H = \{x \in \mathbb{R}^n : \beta^\top x = c\}$ for some vector β and some scalar $c \neq 0$. Let α be such that $\beta^\top \alpha = 0$ and let d be a scalar. Then, there is a vector γ such that*

$$\{x \in H : \alpha^\top x \geq d\} = \{x \in H : \gamma^\top x \geq c\}.$$

Proof. It is easy to check that the following vector does the job

$$\gamma = \begin{cases} c\alpha/d & \text{if } cd > 0, \\ \alpha + \beta & \text{if } d = 0, \\ -c\alpha/d + 2\beta & \text{if } cd < 0. \end{cases} \qquad \square$$

Definition 5. We say that two convex sets $\mathcal{P}_1$ and $\mathcal{P}_2$ *intersect all of the same supporting hyperplanes* if

- they have the same closure, and
- for every supporting hyperplane H, $H \cap \mathcal{P}_1 \neq \emptyset$ if and only if $H \cap \mathcal{P}_2 \neq \emptyset$.

Definition 6. Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be convex sets in $\mathbb{R}^n$. For $i = 1, 2$, define $R_{\mathcal{P}_i}$ as in (3). Let E be a subset of $\mathbb{R}^n$. We say that E *distinguishes* $\mathcal{P}_1$ *and* $\mathcal{P}_2$ if $R_{\mathcal{P}_1}(E) \neq R_{\mathcal{P}_2}(E)$.

We break the proof of Theorem 1 into two parts according to whether or not $\mathcal{P}_1$ and $\mathcal{P}_2$ intersect all of the same supporting hyperplanes. The first part deals with cases in which a single pair of acts can distinguish two convex sets.

Lemma 9. *Suppose that two convex sets $\mathcal{P}_1$ and $\mathcal{P}_2$ do not intersect all of the same supporting hyperplanes. Then there is a set E with one constant act and one possibly nonconstant act that distinguishes $\mathcal{P}_1$ and $\mathcal{P}_2$.*

Proof. First, consider the case in which $\mathcal{P}_1$ and $\mathcal{P}_2$ don't have the same closure. Without loss of generality, let $p_0 \in \mathcal{P}_2 \cap \overline{\mathcal{P}_1}^C$. Let $x \in \mathbb{R}^n$ and c be such that $x^\top p > c$ for all $p \in \overline{\mathcal{P}_1}$ and $x^\top p_0 < c$. Let E consist of the two acts x and the constant $\boldsymbol{c} = (c, \ldots, c)$. Clearly, $\{\boldsymbol{c}\} = R_{\mathcal{P}_1}(E)$ while $\boldsymbol{c} \notin R_{\mathcal{P}_2}(E)$.

Next, consider the case in which $\mathcal{P}_1$ and $\mathcal{P}_2$ have the same closure. Without loss of generality, let $\{p : x^\top p = c\}$ be a supporting hyperplane that intersects $\mathcal{P}_2$ but not $\mathcal{P}_1$ so that $x^\top p > c$ for all $p \in \mathcal{P}_1$. Let $E = \{\boldsymbol{c}, x\}$. Then $\{\boldsymbol{c}\} = R_{\mathcal{P}_1}(E)$ while $\boldsymbol{c} \notin R_{\mathcal{P}_2}(E)$. $\square$

The following result handles the case in which pairwise choice is not sufficient to distinguish two sets. The proof can be summarized as follows. Start with two distinct convex sets of probabilities that intersect all of the same supporting hyperplanes. Find a supporting hyperplane that they intersect in different ways and use this as the first gamble in the set E in such a way that all probabilities in the hyperplane give the gamble the same expected value (say c) and the rest of both convex sets give the gamble smaller expected value. Put the constant $\boldsymbol{c}$ into E as well. Now, the only probabilities that keep the first gamble out of the rejection set are in the hyperplane. We now add further gambles to E in a sequence such that the next one has expected value greater than c except on a boundary of one less dimension than the previous one. By so doing, we reduce the set of probabilities that keep the first gamble out of the rejection set by decreasing its dimension by one each time. Eventually, we get the set of such probabilities to a zero-dimensional set (a single point) that lies in one of the two original convex sets but not the other.

Lemma 10. *Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be distinct convex sets of probabilities in $\mathbb{R}^s$ ($s \geq 2$) that intersect all of the same supporting hyperplanes. Then there is a set E with at most $s+1$ gambles that distinguishes $\mathcal{P}_1$ and $\mathcal{P}_2$.*

Proof. Clearly the difference between $\mathcal{P}_1$ and $\mathcal{P}_2$ is all on the common boundary. Hence, there is some supporting hyperplane that intersects both sets but in different ways. Let such a hyperplane be $H_1 = \{p : x_1^\top p = c\}$ such that for all $p \in \mathcal{P}_i$, $x_1^\top p \leq c$ for $i = 1, 2$. Let $\mathcal{P}_{i,1} = \mathcal{P}_i \cap H_1$ for $i = 1, 2$. Let the first two gambles in E be x_1 and $\boldsymbol{c}$. (If $c = 0$, add a constant to every coordinate of x_1 and replace c by that constant.) The remainder of the proof proceeds through at most $s - 1$ additional steps of the type to follow where one new gamble gets added to E at each step. Initialize $j = 1$.

By construction, $\mathcal{P}_{1,j}$ and $\mathcal{P}_{2,j}$ are distinct convex sets that lie in an $s - j$ dimensional hyperplane. If these sets intersect all of the same supporting hyperplanes (case 1), then find a supporting subhyperplane H'_{j+1} of H_j that intersects $\mathcal{P}_{1,j}$ and $\mathcal{P}_{2,j}$ in different ways. If the sets $\mathcal{P}_{1,j}$ and $\mathcal{P}_{2,j}$ don't intersect all of the same supporting hyperplanes (case 2), use Lemma 9 to find a subhyperplane H'_{j+1} of H_j that distinguishes them. In either case, use Lemma 8 to extend H'_{j+1} to $H_{j+1} = \{p : x_j^\top p = c\}$ such that $x_j^\top p \geq c$ for all $p \in \mathcal{P}_{i,j}$ for both $i = 1, 2$. Include x_j in E. Define $\mathcal{P}_{i,j+1} = \mathcal{P}_{i,j} \cap H_{j+1}$ for $i = 1, 2$.

If case 2 holds in the previous paragraph, skip to the next paragraph. If case 1 holds in the previous paragraph, then increment j to $j + 1$ and repeat the construction in the previous paragraph. Continue in this way either until case 2 holds or we arrive at $j = s - 1$ with one-dimensional sets $\mathcal{P}_{1,s-1}$ and $\mathcal{P}_{2,s-1}$, which then must be bounded line segments. They differ by at least one of them containing a point that the other does not contain. Without loss of generality, suppose that $\mathcal{P}_{2,s-1}$ contains a point p_0 that is not in $\mathcal{P}_{1,s-1}$. Create one last vector x_s so that $x_s^T p_0 = c$ and $x_s^\top p > c$ for all $p \in \mathcal{P}_{1,s-1}$.

Every gamble $x \in E$ satisfies $x^\top p_0 = c$, while for every $p \in \mathcal{P}_1$, there is $k \geq 2$ such that $x_k^\top p > c$. It now follows that $x_1 \in R_{\mathcal{P}_1}(E)$ but $x_1 \notin R_{\mathcal{P}_2}(E)$. $\square$

References

Rubin, H. (1987). A weak system of axioms for "rational" behavior and the non-separability of utility from prior. *Statistics & Decisions*, **5**, 47–58.

Schervish, M. J., Seidenfeld, T., Kadane, J. B., and Levi, I. (2003). Extentions of expected utility theory and some limitations of pairwise comparisons. In *Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications* (J.-M. Bernard, T. Seidenfeld and M. Zaffalon, eds.) 496–510. Carlelton Scientific.

Seidenfeld, T., Schervish, M. J., and Kadane, J. (1995). A representation of partially ordered preferences. *Annals of Statistics*, **23**, 2168–2217. MR886877

Sen, A. (1977). Social choice theory: A re-examination. *Econometrica*, **45**, 53–89. MR456612

Von Neumann, J. and Morgenstern, O. (1947). *Theory of Games and Economic Behavior* (2nd ed.) Princeton University Press, Princeton. MR21298

Walley, P. (1990). *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall, London. MR1145491