Information, Subjective Belief and Preference

Takashi Hayashi*
Department of Economics, University of Rochester

October 2003

Abstract

This paper presents an axiomatic model of decision making which incorporates objective but imprecise information as a variable. *The model achieves two primary objectives. First, it permits the analyst to relate the choices made under different information. Second, it explains how subjective belief varies with objective information.* In contrast, note that in the von Neumann-Morgenstern expected utility model information is objective and precise, and that in Savage's subjective expected utility theory and in the multiple-priors model due to Gilboa-Schmeidler, information is implicit and fixed. This leaves subjective priors unexplained and limits the ability to calibrate the model, that is, to relate behavior of the single DM in different settings.

We adopt a Savage-style state space model. Information is assumed to take the form of a probability-possibility set, that is, a set P of probability measures on the state space. The DM is told only that the true probability law lies in P. She is assumed to rank pairs of the form (P, f) where P is a probability-possibility set and f is an act mapping states into outcomes.

The representation result delivers multiple-priors utility at *each* probability-possibility set. The subjective set of priors is obtained by (i) solving for the 'mean value' of the probability-possibility set, and (ii) shrinking the probability-possibility set toward the mean value to a degree determined by preference. This allows both subjective expected utility with the prior selected as the mean value, and the other extreme where the DM takes the worst case scenario in the entire probability-possibility set. The degree of shrinkage corresponds to the DM's degree of 'imprecision aversion,' and it can be elicited by a simple procedure.

*I am grateful to Larry Epstein for his patient encouragement and invaluable advice, and to Mark Machina, Faruk Gul and Massimo Marinacci for helpful comments.

1 Introduction

1.1 Objectives

Consider a decision maker (DM) who has multiple-priors preferences as in Gilboa-Schmeidler [8]. To evaluate an act she computes the minimum value of expected utility as the probability measure varies over her subjective set of priors. The DM faces two alternative Ellsberg urns. Urn 1 contains 100 balls that are red, green or blue. She is told that there are 40 red balls but she is told nothing about the numbers of green and blue balls. Suppose we observe the DM's choices between bets on this urn. Now consider the same DM confronted with a second urn containing 100 balls with the same possible colors, but where she is told that the number of red balls is between 40 and 80, the number of green balls is between 10 and 50, and nothing about the number of blue balls.

The following two questions motivate this paper:

1. *Given the DM's behavior in urn 1, can we predict her behavior in the second urn?*

2. *How does the subjective set of priors vary with the objective information?*

Consider the two principal theories of choice under uncertainty - the von Neumann-Morgenstern expected utility theory [19] and the subjective expected utility theory due to Savage [14] and Anscombe-Aumann [2]. Do they provide an answer?

In the vNM theory, the DM is given a lottery, a probability distribution over outcomes, as an object of choice. Thus, information about consequence of any given choice is *objective* and *precise.* This is not the case above, since the DM is not told the true proportions of the colors in either urn. In subjective expected utility theory, the DM chooses between acts, random variables mapping states of the world into outcomes. Information about states is *implicit* and *fixed.* Therefore, the model cannot be used to relate behaviors in two situations such as when facing two alternative Ellsberg urns. A similar remark applies to Schmeidler [15] and Gilboa-Schmeidler [8]. Their models accommodate ambiguity aversion, but do not relate preferences across two situations.

We present an axiomatic model of decision making which incorporates objective but imprecise information as a variable. The model permits the analyst to relate the choices made under different information. Further, it explains how subjective belief is related to objective information.

We adopt a Savage-style state space model. Information is assumed to take the form of a set P of probability measures on the state space - the DM is told only that the true probability law lies in P. We call such a set a *probability-possibility set*. In the Ellsberg example above, urn 1 is represented by $P_1 = \{(p_R, p_G, p_B) : p_R = 0.4\}$, and urn 2 is represented by $P_2 = \{(p_R, p_G, p_B) : 0.4 \leq p_R \leq 0.8,\ 0.1 \leq p_G \leq 0.5\}$.

The DM is assumed to rank pairs each consisting of a probability-possibility set and an act; as in Savage, an act maps states into outcomes. Let P and Q denote probability-possibility sets and let f and g denote acts. When she prefers taking f under P to taking g under Q, we write

$$(P, f) \succsim (Q, g).$$

When $Q = P$, the above preference relation represents the ranking of acts given the information embodied by P. The Savage model may be viewed as this case with some fixed P. By varying P, we obtain the ranking of acts conditional on different information. When $g = f$, the preference relation represents the ranking of information given the action embodied by f. Thus, by varying f we can look at the DM's attitude toward information given an act. Finally, the model can be viewed as a model of risk preference when information is precise, namely when the probability-possibility set is a singleton.

1.2 Overview of results

The representation result delivers multiple-priors utility over acts conditional on *each* probability-possibility set. The subjective set of priors is obtained by (i) solving for the 'mean value' of the probability-possibility set, and (ii) shrinking the probability-possibility set toward the mean value according to a degree given by preference (see Figure 1). The 'mean value' is the *Steiner point* (see [17]). It is roughly the weighted average of extreme points of the probability-possibility set in which the weight for each is proportional to its outer angle.

To illustrate, suppose the probability-possibility set is given by the triangle ABC in Figure 0 where the outer angle is 120° at A, 135° at B and 105° at C. The Steiner point is computed by $\frac{120}{360}A + \frac{135}{360}B + \frac{105}{360}C$.

For any probability-possibility set/act pair (P, f), its utility is expressed in the form

$$U(P, f) = \min_{p \in \varphi(P)} E_p \left[u \circ f \right] \qquad (*)$$

and

$$\varphi(P) = (1 - \varepsilon)\{s(P)\} + \varepsilon\, P, \qquad (**)$$

where $\varepsilon \in [0, 1]$ and $s(P)$ denotes the Steiner point of P.[1]

[1] $E_p[\,\cdot\,]$ denotes expectation with respect to p.

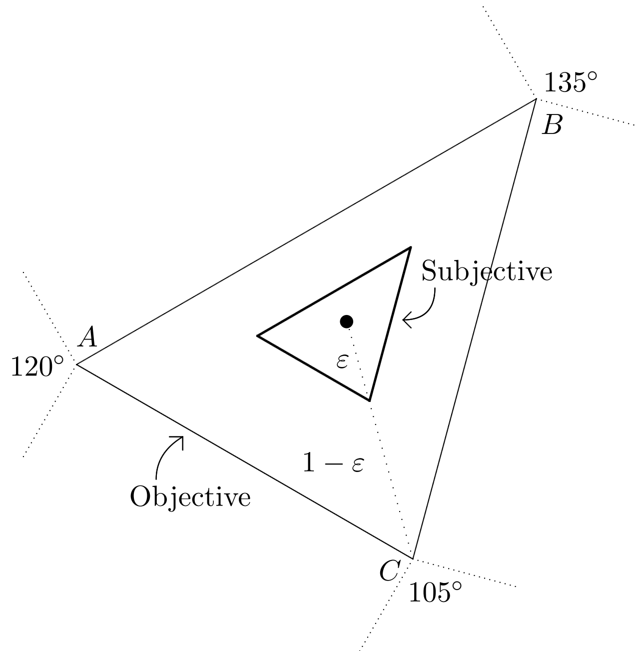


Figure 1: Objective information and subjective set of priors.

The parameter ε is obtained as a part of the representation result and can be interpreted as a (subjective) degree of 'imprecision aversion.' When $\varepsilon = 0$, we obtain subjective expected utility in which the prior is the Steiner point. When $\varepsilon = 1$, the functional form expresses the extreme case where the DM takes the worst case scenario in the entire probability-possibility set.

The degree of imprecision aversion ε can be elicited by a simple procedure. Bring two urns, each containing 100 balls which are possibly red or blue. The DM is told nothing about the proportions in urn 1. For urn 2, she is told that the proportion of red is p. Denote a bet on red in urn i by R_i, where the bet yields payoff x^* if red is drawn from i, and x if not, $x \prec x^*$. Suppose p is such that the DM is indifferent between R_1 and R_2. Such p may be called a 'precise information equivalent.' Then, the DM's degree of imprecision aversion is given by $\varepsilon = 1 - 2p$. When the DM faces another probability-possibility set P, her subjective set of priors is $(1-\varepsilon)\{s(P)\}+\varepsilon P$. Since $u(\cdot)$ can be elicited through choice between lotteries, the model permits prediction of behavior under different information.

Return to the example in the beginning. The subjective set of priors for each probability-possibility set is given by $\varphi(P_1) = (1-\varepsilon)\{(0.4, 0.3, 0.3)\} + \varepsilon\{p : p_R = 0.4\}$ and $\varphi(P_2) = (1-\varepsilon)\{(0.55, 0.25, 0.2)\} + \varepsilon\{p : 0.4 \leq p_R \leq 0.8,\ 0.1 \leq p_G \leq 0.5\}$.

1.3 Second-order prior model

An alternative Bayesian-style model of preference is one where the DM has a 'second-order' prior over probability measures and updates it conditional on the given probability-possibility set. It may be formulated as

$$U(P,f) = \begin{cases} \int_P \phi\left(E_p\left[u \circ f\right]\right)\ \nu(dp)/\nu(P) & \text{if} \qquad \nu(P) > 0, \\ \phi\left(E_p\left[u \circ f\right]\right) & \text{if} \qquad P = \{p\}, \end{cases} \tag{$\star$}$$

where ν is a (non-atomic) probability measure over the probability simplex and where ϕ is a function which expresses the DM's attitude towards imprecision. When ϕ is concave, she is averse to imprecision. When ϕ is linear, the functional form reduces to

$$U(P,f) = E_{\nu_P}\left[u \circ f\right],$$

where

$$\nu_P = \begin{cases} \int_P p\,\nu(dp)/\nu(P) & \text{if} \qquad \nu(P) > 0, \\ p & \text{if} \qquad P = \{p\}. \end{cases}$$

This is subjective expected utility with prior ν_P. When ν is uniform, ν_P is a *conditional mean* of P. Conditional mean and the Steiner point are distinct notions of mean value for sets. As a result, even the Bayesian special cases of the two models, $\varepsilon = 0$ in ($**$) and ϕ linear in ($\star$), are distinct.

The relation between our model and the second-order prior model will be discussed further below.

1.4 Outline and related literature

The paper proceeds as follows. In the remainder of the introduction, we summarize the related literature. Section 2 introduces the primitives of the model. Section 3 presents the axioms. In section 4, we present the main representation result. In section 5, we show how the parameter ε represents the DM's imprecision aversion. Proofs are collected in the appendix.

Gajdos, Tallon and Vergnaud [9] assume that the DM chooses between triples consisting of a 'base prior', a probability-possibility set and an act. They characterize multiple-priors utility in which the subjective set of priors is obtained by shrinking the probability-possibility set toward the base prior to a degree determined by preference. In our model base priors are not taken to be objective and observable by the analyst. Rather, they are subjective and delivered as a part of the representation of preference. Wang [20] adopts the same domain as in [9] and characterizes the extreme case where the DM takes the worst case scenario in the entire probability-possibility set.

Damiano [6] studies a family of preferences over acts indexed by probability-possibility sets, $\{\succsim_P\}$, where P is given as the core of a lower probability. He characterizes a version of subjective expected utility in which the prior is selected as the nucleolus of lower probability.

Another way of formulating objective but imprecise information is by adopting sets of lotteries over outcomes as the objects of choice. In this approach, Olszewski [13] characterizes a version of α-maxmin expected utility in which the DM puts weights both on the best-case and the worst-case scenarios. Ahn [1] characterizes a conditional subjective expected utility in which the DM has a prior probability over lotteries, and updates it according to each objective set. The relation between their models and ours will be discussed further below.

Jaffray [11] studies a model in which capacities over outcomes are the objects of choice. He characterizes a non-additive version of vNM expected utility.

2 Primitives

The set of possible states of the world, denoted Ω, is assumed to be finite and $|\Omega| \geq 3$.[2] Denote the set of probabilities over Ω by $\Delta(\Omega)$ and its typical element by p. Denote by $\mathcal{P}$ the space of *non-empty compact convex* subsets of $\Delta(\Omega)$, endowed with the Hausdorff metric; then $\mathcal{P}$ is compact (see pp. 49-50 in [17]). We call an element of $\mathcal{P}$, typically denoted by P, a *probability-possibility set*. When told P, the DM is assumed to know only that the true probability lies in P. When a probability-possibility set is given as a singleton typically denoted by $\{p\}$, the DM knows the true probability precisely and we say there is *precise information*.

The space of probability-possibility sets $\mathcal{P}$ is a mixture space under the operation defined by

$$\lambda P + (1-\lambda)Q = \{\lambda p + (1-\lambda)q : p \in P, q \in Q\}.$$

Modelling information by sets of probabilities is restrictive; for example, it does not permit distinguishing between two situations where the sets of possible probability laws are the same but where some probabilities are more likely in one situation than in the other. Ellsberg [7] provides an example where information *does* come in the form of a convex set of probabilities.

[2]Throughout the paper, we fix the state space Ω. One may view our model as a variable state space model by saying that Ω is the set of 'potential' states of the world, and that $S \subset \Omega$ is the present state space when probabilities of states outside of S are known to be zero. Also one can treat a two-state (one-state) case by saying that probabilities of states other than the two (one) are zero. However, our analysis requires that there are at least three potential states of the world.

The outcome space X is assumed to be compact metric. Let $\Delta(X)$ be the set of lotteries (Borel measures) over X, endowed with the weak convergence topology. It is again compact metric with respect to the Prokhorov metric. Let $\mathcal{F} = \{f : \Omega \to \Delta(X)\}$ be the set of *lottery acts*.

The domain of objects of choice is $\mathcal{P} \times \mathcal{F}$. It is endowed with the product topology. The DM has a preference relation over $\mathcal{P} \times \mathcal{F}$, which is denoted by $\succsim$. The DM compares pairs of probability-possibility sets and acts. When

$$(P, f) \succsim (Q, g),$$

the DM prefers choosing f under P to choosing g under Q. When $Q = P$, the preference relation represents the ranking of acts given the information embodied by P. When $g = f$, the preference relation represents the ranking of probability-possibility sets given the action embodied by f.

3 Axioms

First, we assume that the DM's preference is a weak order.

Axiom 1 (Order): The preference relation $\succsim$ is complete and transitive.

Second, we assume that preference is continuous.

Axiom 2 (Continuity): The preference relation $\succsim$ is continuous on $\mathcal{P} \times \mathcal{F}$.

In particular, preference is continuous over the space of probability-possibility sets with respect to the Hausdorff metric. It is not innocuous when non-convex sets are treated. Consider a three-state case, and take two pp sets $P = \{(1, 0, 0), (0, 1, 0)\}$ and $Q = \{(1, 0, 0), (0, 1, 0), (0.01, 0.99, 0), (0, 0.99, 0.01)\}$ (see Figure 2). In the Hausdorff metric, they are very 'close.' However, it may be plausible that the DM views state 2 as much more likely given Q than in P and thus rank them very differently. When the probability-possibility set is given as a convex set, any mixture of its elements is also regarded to be possible (see Figure 3). We avoid the above kind of phenomena by restricting attention to convex sets.

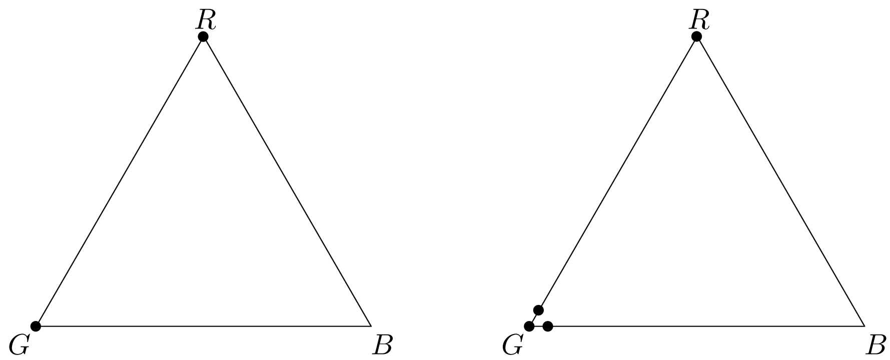


Figure 2: Non-convex probability-possibility sets

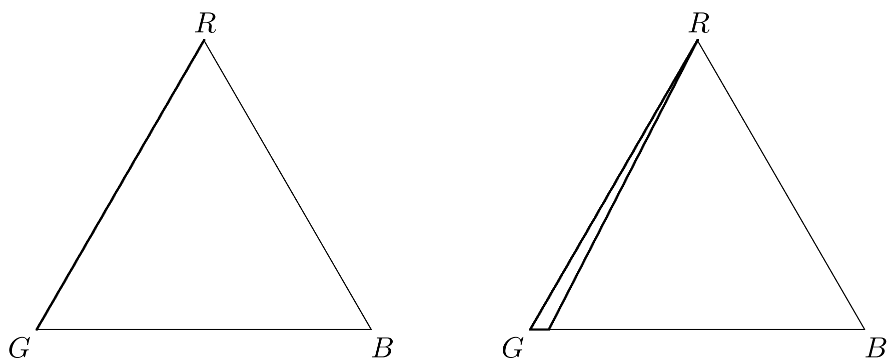


Figure 3: Convex probability-possibility set

Axiom 3 (Outcome Preference): (i) For any $P, P' \in \mathcal{P}$ and $l \in \Delta(X)$, $(P, l) \sim (P', l)$, and (ii) there exist $P \in \mathcal{P}$ and $l, l' \in \Delta(X)$ such that $(P, l) \succ (P, l')$.

The third axiom states that the preference over lotteries is independent of probability-possibility sets and is nondegenerate. When a lottery is given regardless of states of the world, information about their likelihood is irrelevant. Also, we exclude the case that the DM is indifferent between all the lotteries.

For probability $p \in \Delta(\Omega)$ and act $f \in \mathcal{F}$, define the induced distribution over outcomes by

$$l(p, f) = \sum_{\omega \in \Omega} p(\omega) f(\omega)$$

Axiom 4 (Reduction under Precise Information): For any $p \in \Delta(\Omega)$ and $f \in \mathcal{F}$,

$$(\{p\}, f) \sim (\{p\}, l(p, f)).$$

This axiom states that the evaluation of an act under *precise* information depends only on its induced distribution. We do not assume the counterpart of this for general probability-possibility sets. As is discussed later, two probability-possibility set/act pairs may be differently evaluated even if they induce the same sets of distributions over outcomes.

Axiom 5 (Independence): For any $P, P', Q \in \mathcal{P}$, $f \in \mathcal{F}$, and $\lambda \in (0,1)$,

$$(P, f) \succ (P', f) \Longrightarrow (\lambda P + (1-\lambda)Q, f) \succ (\lambda P' + (1-\lambda)Q, f).$$

To interpret, consider the formal object '$\lambda \circ P + (1-\lambda) \circ Q$'. Given this, the DM knows that 'P is true with probability λ and Q is true with probability $1-\lambda$.' Suppose the DM is to evaluate the two objects $\lambda \circ P + (1-\lambda) \circ Q$ and $\lambda P' \circ + (1-\lambda) \circ Q$ in choosing an act. Then, the difference between them is only in P and P' which are true 'with probability λ'. It is intuitive that the ranking of these should depend only on the ranking of P and P', and the common probability-possibility Q being true 'with probability $1-\lambda$' should not matter.

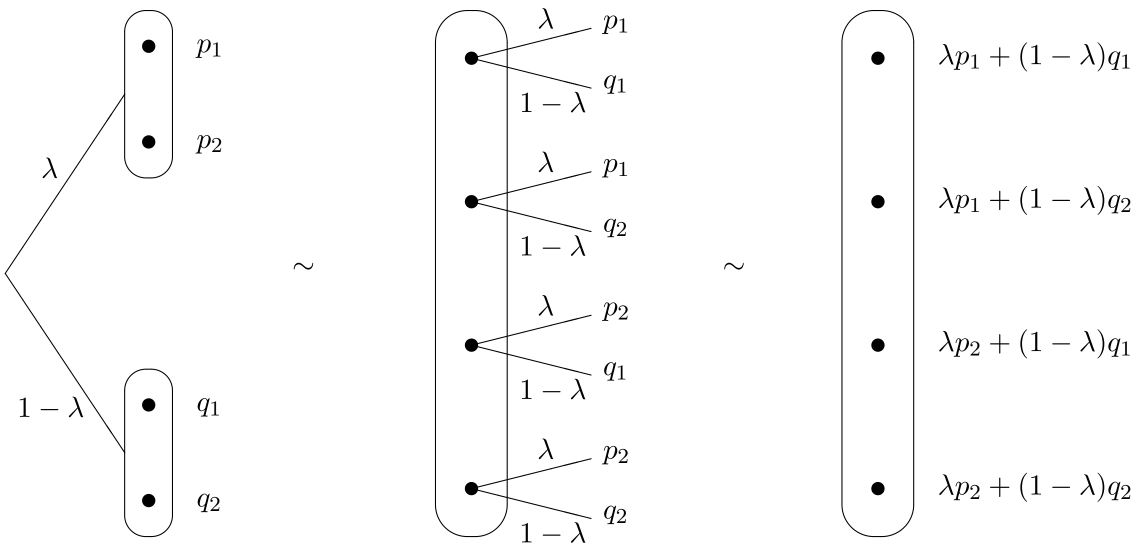
We explain why the axiom permits the above interpretation, by arguing how the DM identifies the object $\lambda \circ P + (1-\lambda) \circ Q$ with $\lambda P + (1-\lambda)Q$. The argument consists of two steps.

First, the DM is indifferent between $\lambda \circ P + (1-\lambda) \circ Q$ and '$\{\lambda \circ p + (1-\lambda) \circ q : p \in P,\ q \in Q\}$.' When the latter is given, the DM knows it is possible that p is true with probability λ and q is true with probability $1-\lambda$, for each $p \in P$ and $q \in Q$.

There are two kinds of uncertainty here. One is about outcome of randomization, which is risk, and the other is about ultimate realization of true probability law. We assume that the DM is indifferent in the order of these two uncertainties (see the left half of Figure 4).

Second, compare $\{\lambda \circ p + (1-\lambda) \circ q : p \in P,\ q \in Q\}$ and $\lambda P + (1-\lambda)Q$. The former is the set of compound probabilities, and the latter is that of their reduced ones. We assume that the DM is indifferent in the reduction of compound probabilities. That is, she is indifferent in the timing of resolution of risk, which is assumed in the standard theory (see the right half of Figure 4).

Thus, we deduce that the DM views three objects $\lambda \circ P + (1-\lambda) \circ Q$, $\{\lambda \circ p + (1-\lambda) \circ q : p \in P,\ q \in Q\}$ and $\lambda P + (1-\lambda)Q$ to be the same, and that the axiom allows the interpretation discussed above.



$\lambda \circ P + (1-\lambda) \circ Q$ $\quad \{\lambda \circ p + (1-\lambda) \circ q : p \in P, q \in Q\}$ $\quad \lambda P + (1-\lambda)Q$

Figure 4: Independence

Axiom 6 (Set Dominance): For any $f \in \mathcal{F}$ and $p \in \Delta(\Omega)$ and $P \in \mathcal{P}$,

$$(\{p'\}, f) \succsim (\{p\}, f) \text{ for every } p' \in P \Longrightarrow (P, f) \succsim (\{p\}, f),$$

and

$$(\{p\}, f) \succsim (\{p'\}, f) \text{ for every } p' \in P \Longrightarrow (\{p\}, f) \succsim (P, f).$$

Fix an act and pick a probability-possibility set and a point. If every element of the set is better (worse) than the point, it will be natural to conclude that the whole set is better (worse) than the point.

Axiom 7 (Act Dominance): For any $f, g \in \mathcal{F}$ and $P \in \mathcal{P}$,

$$(\{p\}, f) \succsim (\{p\}, g) \text{ for every } p \in P \Longrightarrow (P, f) \succsim (P, g).$$

If one act is preferable to another under every element of the probability-possibility set, it will be natural to conclude that the ranking is unchanged even under the whole set.

Axiom 8 (Gains via Hedging): For any $f, g \in \mathcal{F}$, and $p \in \Delta(\Omega)$, $P \in \mathcal{P}$,

$$(P, f) \sim (\{p\}, f) \text{ and } (P, g) \sim (\{p\}, g) \Longrightarrow (P, \lambda f + (1-\lambda)g) \succsim (\{p\}, \lambda f + (1-\lambda)g).$$

The precise information $\{p\}$ here will be seen as a 'precise information equivalent' of the imprecise information P, which depends on acts. Suppose p is a precise information equivalent of P with respect to each of acts f and g. When the DM mixes acts, it reduces variation across states. This renders imprecision of information less costly, whereas it is neutral to precise information. Thus, the imprecise information will be evaluated better than its precise information equivalent for the original choices.

Axiom 9 (Cardinal Invariance): For any $f \in \mathcal{F}$, $l \in \Delta(X)$, $\lambda \in (0, 1)$ and $P, Q \in \mathcal{P}$,

$$(P, f) \succsim (Q, f) \text{ if and only if } (P, \lambda f + (1-\lambda)l) \succsim (Q, \lambda f + (1-\lambda)l).$$

To illustrate, suppose $(\{p\}, f) \succsim (\{q\}, f)$ for precise information $p, q \in \Delta(\Omega)$ and act f. When a constant act $l \in \Delta(X)$ is given, information about likelihood of states is irrelevant (Axiom 2). That is, we have $(\{p\}, l) \sim (\{q\}, l)$. Under axioms 1-5, the preference accommodates to the expected utility structure when information is precise. Thus, the evaluation of precise information is unaffected by mixing with a constant act, which leads to the ranking $(\{p\}, \lambda f + (1-\lambda)l) \succsim (\{q\}, \lambda f + (1-\lambda)l)$.

The axiom states that such invariance is valid also for general probability-possibility sets. Since the ranking of any probability is unchanged by mixing with constant act, it is intuitive that even the ranking of probability-possibility sets is unchanged.

The last axiom we are going to introduce states that the DM's attitude toward information should not change in the transformation of state space which does not change its substance. Let $\psi : \Delta(\Omega) \to \Delta(\Omega)$ be such a transformation, and let $\widetilde{\psi} : \mathcal{F} \to \mathcal{F}$ be the transformation of act associated with ψ. Then, the axiom takes the form that for any $P, Q \in \mathcal{P}$ and $f \in \mathcal{F}$,

$$(P, f) \succsim (Q, f) \Longrightarrow (\psi(P), \widetilde{\psi}(f)) \succsim (\psi(Q), \widetilde{\psi}(f)).$$

Now the problem is what class of transformations should be considered for such ψ.

The simplest one is permutation of state space. Consider a permutation $\pi : \Omega \to \Omega$. Then, the permutation of probability-possibility set is defined by $\pi P = \{q \in \Delta(\Omega) : q(\omega) = p(\pi^{-1}(\omega)) \text{ for each } \omega \in \Omega,\ p \in P\}$, and the permutation of act πf is defined by $(\pi f)(\omega) = f(\pi^{-1}(\omega))$ for each $\omega \in \Omega$.

To illustrate, take the three-color urn example (see Figure 4) in which $\Omega = \{R, G, B\}$. Consider f given by $f(R) = 100$, $f(G) = 0$ and $f(B) = 0$. Suppose first that given f the DM prefers P to Q. Now rename balls as red to green, green to blue, and blue to red. That is, permutation π is made according to the graph in Figure 5. Then, πf is given by $(\pi f)(R) = 0$, $(\pi f)(G) = 100$ and $(\pi f)(B) = 0$, which means betting on $\Pi R = G$. It is intuitive that given πf the DM prefers πP to πQ (see Figure 6).

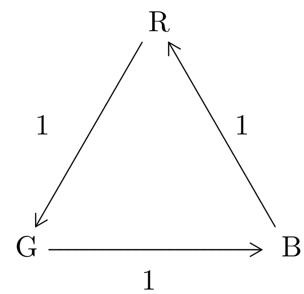


Figure 5: Permutation

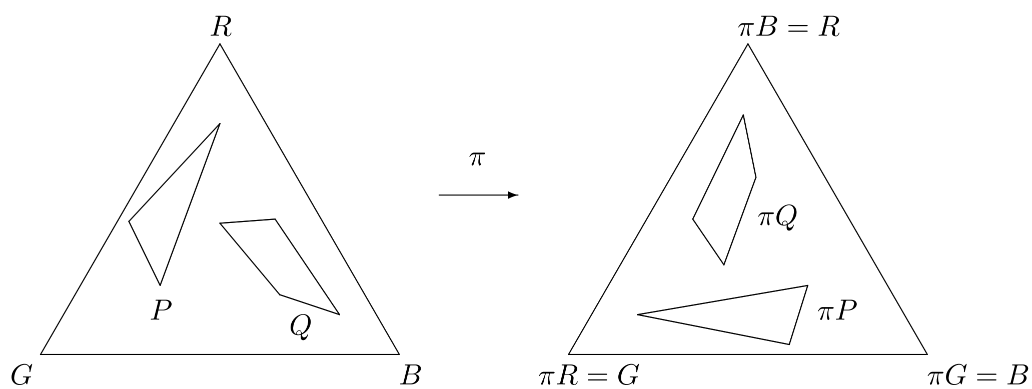


Figure 6: Invariance to permutation

Thus, permutation is the natural candidate of transformation in which the DM's attitude toward information should be unchanged. Next, we consider a broader class which consists of their 'stochastic extensions'. An $|\Omega| \times |\Omega|$-matrix Π is *bistochastic* if it is nonnegative and $\sum_{\omega \in \Omega} \Pi_{\omega\omega'} = 1$ for each $\omega' \in \Omega$, and $\sum_{\omega' \in \Omega} \Pi_{\omega\omega'} = 1$ for each $\omega \in \Omega$. For a bistochastic matrix Π and $f \in \mathcal{F}$, define the transformed act $\Pi f \in \mathcal{F}$ by $(\Pi f)(\omega) = \sum_{\omega' \in \Omega} \Pi_{\omega\omega'} f(\omega')$ for each $\omega \in \Omega$.

Any bistochastic matrix may be expressed as a convex combination of permutation matrices (see [3]). In that sense, it is a stochastic generalization of permutation. However, a bistochastic matrix can change the substance of state space. To clarify, consider a transformation by a bistochastic matrix according to the graph in Figure 7. Consider f given by $f(R) = 100$, $f(G) = 0$ and $f(B) = 0$. Given f, state R is best and G and B are the worst cases. Thus, $(\{R\}, f) \succ$

$(\{G\}, f) \sim (\{B\}, f)$. Now transform the state space, to obtain $\Pi R = \left(\frac{2}{3}, \frac{1}{3}, 0\right)$, $\Pi G = \left(\frac{1}{3}, \frac{2}{3}, 0\right)$ and $\Pi B = (0, 0, 1)$ (see Figure 8). The transformed act Πf delivers $(\Pi f)(R) = \left(100; \frac{2}{3}, 0; \frac{1}{3}\right)$, $(\Pi f)(G) = \left(100; \frac{1}{3}, 0; \frac{2}{3}\right)$ and $(\Pi f)(B) = 0$. View $\{\Pi R, \Pi G, \Pi B\}$ as a 'new state space.' Then Πf can be seen as an act that maps $\{\Pi R, \Pi G, \Pi B\}$ to outcomes, and delivers $(\Pi f)(\Pi R) = \left(100; \frac{5}{9}, 0; \frac{4}{9}\right)$, $(\Pi f)(\Pi G) = \left(100; \frac{4}{9}, 0; \frac{5}{9}\right)$ and $(\Pi f)(\Pi B) = 0$. Thus Πf delivers a better outcome in 'state' ΠG than in 'state' ΠB; that is, $(\{\Pi G\}, \Pi f) \succ (\{\Pi B\}, \Pi f)$ even though $(\{G\}, f) \sim (\{B\}, f)$. Since even the ranking of precise probabilities is changed by the transformation, the DM does not view the new probability simplex as substantially the same as the original one.

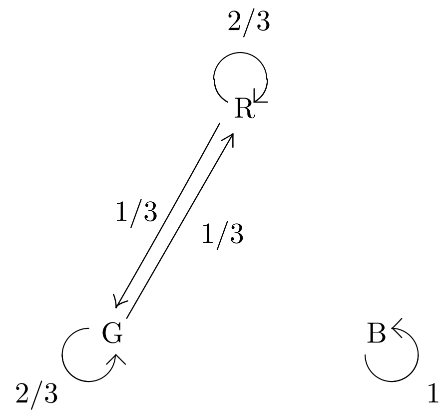


Figure 7: Transformation by a bistochastic matrix.

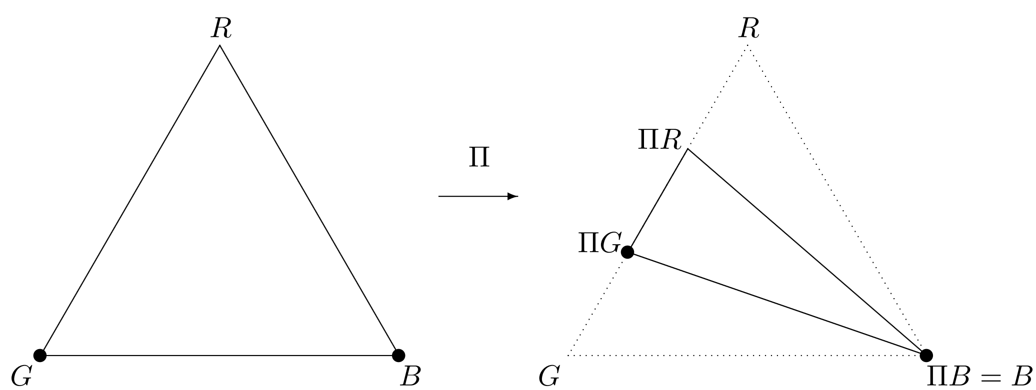


Figure 8: Transformation of state space by a bistochastic matrix.

We formulate the invariance axiom using a subclass of bistochastic matrices. A matrix Π is a *similarity* if there exists $\lambda \geq 0$ such that $\|\Pi p - \Pi q\| = \lambda \|p - q\|$ for any $p, q \in \Delta(\Omega)$. When a matrix is bistochastic and is a similarity, we call it a *similarity reshuffle*. Let SR be the set of similarity reshuffles. Note that the image of probability simplex by a similarity reshuffle is similar to the original one in the sense of Euclidian geometry. It follows that the transformation in the previous example is not a similarity, the simplex defined by $\{\Pi R, \Pi G, \Pi B\}$ is not similar to the original one.

To illustrate, consider a similarity reshuffle according to the graph in Figure 9. The probability simplex is transformed into a new probability simplex spanned by ΠR, ΠG and ΠB which are viewed as new states of the world. The new probability simplex is similar to the original one (see Figure 10).

There is a behavioral counterpart to the above geometric similarity. Similarity reshuffles preserve the ranking of probabilities induced by acts. Precisely, we have

Lemma 1 *Assume axioms 1-5. For any* $\Pi \in SR$, $p, q \in \Delta(\Omega)$ *and* $f \in \mathcal{F}$, $(\{p\}, f) \succsim (\{q\}, f)$ *implies* $(\{\Pi p\}, \Pi f) \succsim (\{\Pi q\}, \Pi f)$.

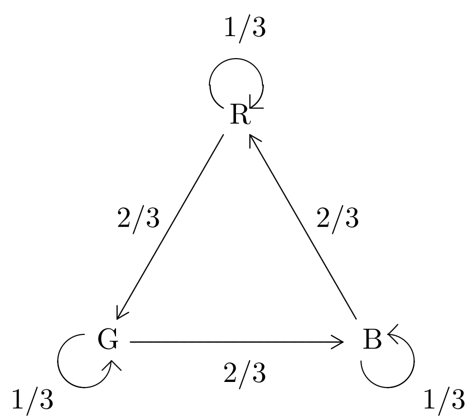


Figure 9: Similarity reshuffle

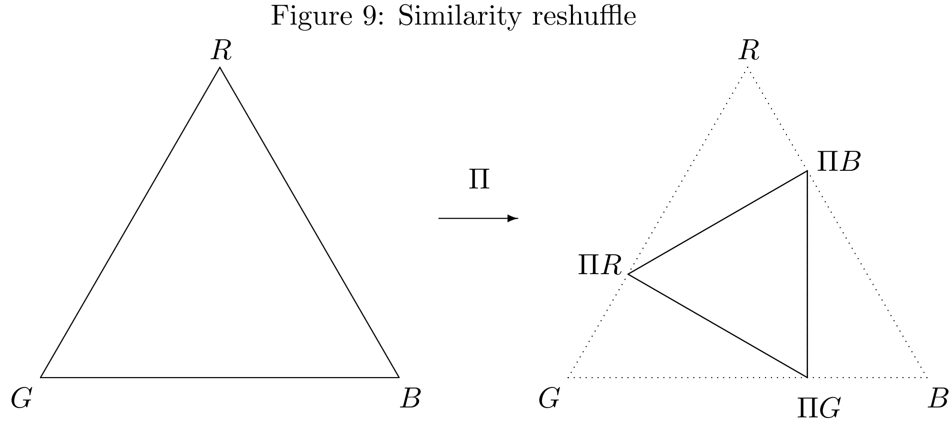


Figure 10: New state space and probability simplex given by similarity reshuffle

Invariance to similarity reshuffles is *implied* by Axioms 1-5 when information is precise. The sought-after final axiom imposes invariance also when information is imprecise.

Axiom 10 (Invariance to Similarity Reshuffles): For any $\Pi \in SR$, $f \in \mathcal{F}$ and $P, Q \in \mathcal{P}$,

$$(P, f) \succsim (Q, f) \Longrightarrow (\Pi P, \Pi f) \succsim (\Pi Q, \Pi f).$$

To interpret, suppose that the DM prefers P to Q given f. Then under axioms 1-5, Πf induces the same ranking of probabilities in the new simplex as f does in the original one. That is, for any $p \in P$ and $q \in Q$,

$$(\{p\}, f) \succsim (\{q\}, f) \Longrightarrow (\{\Pi p\}, \Pi f) \succsim (\{\Pi q\}, \Pi f),$$

and

$$(\{q\}, f) \succsim (\{p\}, f) \Longrightarrow (\{\Pi q\}, \Pi f) \succsim (\{\Pi p\}, \Pi f).$$

Thus in the new probability simplex when given Πf, ΠP and ΠQ play the same roles as P and Q do in the original one when given f. Therefore, it is intuitive that the ranking of possibility sets is unchanged, which leads to the ranking $(\Pi P, \Pi f) \succsim (\Pi Q, \Pi f)$.

To illustrate, take the three-color urn example again and consider f given by $f(R) = 100$, $f(G) = 0$ and $f(B) = 0$. Given f, state R is best and G and B are the worst cases. Suppose first that given f the DM prefers P to Q. Now reshuffle the names of the balls according the graph in Figure 10. Then Πf delivers $(\Pi f)(R) = \left(100; \frac{1}{3}, 0; \frac{2}{3}\right)$, $(\Pi f)(G) = \left(100; \frac{2}{3}, 0; \frac{1}{3}\right)$ and $(\Pi f)(B) = 0$. The similarity reshuffle Π gives the new state space $\{\Pi R, \Pi G, \Pi B\}$, and the reshuffled act Πf is viewed as an act which maps $\{\Pi R, \Pi G, \Pi B\}$ to outcomes and delivers $(\Pi f)(\Pi R) = \left(100; \frac{5}{9}, 0; \frac{4}{9}\right)$, $(\Pi f)(\Pi G) = \left(100; \frac{2}{9}, 0; \frac{7}{9}\right)$ and $(\Pi f)(\Pi B) = \left(100; \frac{2}{9}, 0; \frac{7}{9}\right)$. Thus given Πf, the new state ΠR is best, and ΠG and ΠB are the worst cases. Since the ranking of any probability is unchanged in the same way, it is intuitive that also the ranking of probability-possibility sets is unchanged.

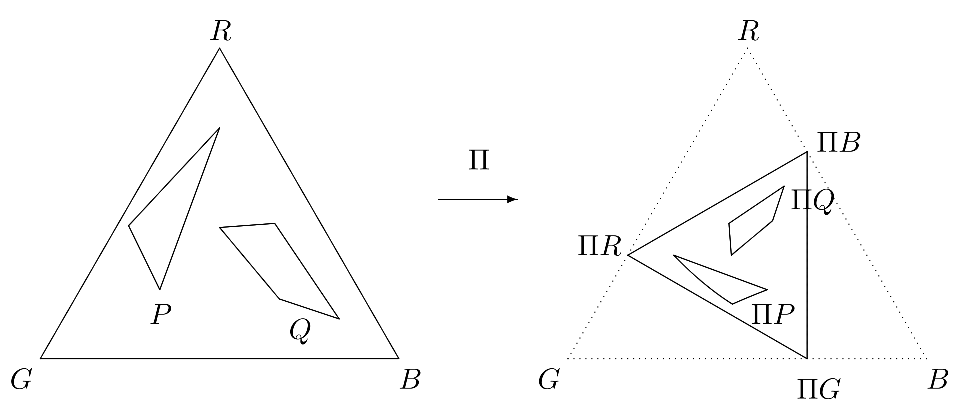


Figure 11: Invariance to similarity reshuffle

4 Representation Theorem

The representation result delivers multiple-priors utility at *each* probability-possibility set. The subjective set of priors is obtained by (i) solving for the 'mean value' of the probability-possibility set, and (ii) shrinking the probability-possibility set toward the mean value according to a degree given by preference. The 'mean value' is the *Steiner point*. Let $e = (\frac{1}{|\Omega|}, \cdots, \frac{1}{|\Omega|})$ and $V = \{v \in R^\Omega : (v-e)e = 0, \|v-e\| = 1\}$ be the $|\Omega| - 2$ dimensional unit sphere around e. For a possibility set $P \in \mathcal{P}$, its Steiner point is defined by

$$s(P) = e + \frac{1}{(|\Omega|-1)\mu(V)} \int_V (v-e)h_P(v-e)\ \mu(dv)$$

where h_P is the support function given by $h_P(v-e) = \max_{p \in P}(p-e)(v-e)$ and μ is the spherical Lebesgue measure over V.

Example 1 *Steiner point of a segment is its midpoint.*

Example 2 *Steiner point of a polytope is the weighted average of its vertices, in which the weight for each vertex is proportional to its outer angle.*

Example 3 *When a probability-possibility set is given as the core of a lower probability (convex capacity), its Steiner point coincides with the Shapley value of the lower probability.*

We state the main theorem.

Theorem 1 *The preference relation $\succsim$ satisfies Axioms 1-10 if and only if there exists a function $U : \mathcal{P} \times \mathcal{F} \to R$ which represents $\succsim$ and a mixture-linear and continuous function $u : \Delta(X) \to R$ and a number $\varepsilon \in [0,1]$ such that*

$$U(P, f) = \min_{p \in \varphi(P)} \sum_{\omega \in \Omega} u(f(\omega))\ p(\omega) \tag{$*$}$$

where

$$\varphi(P) = (1-\varepsilon)\{s(P)\} + \varepsilon\, P. \tag{$**$}$$

Moreover, u is unique up to positive linear transformations and ε is unique.

An immediate corollary is obtained by strengthening Axiom 8 to

Axiom 8* (Hedging Neutrality): For any $f, g \in \mathcal{F}$, and $p \in \Delta(\Omega)$, $P \in \mathcal{P}$,

$(P, f) \sim (\{p\}, f)$ and $(P, g) \sim (\{p\}, g)$

imply $(P, \lambda f + (1-\lambda)g) \sim (\{p\}, \lambda f + (1-\lambda)g)$.

That is, hedging is neutral to the evaluation of possibility sets.

Corollary 1 *The preference relation $\succsim$ satisfies Axioms 1-7,8*,9,10 if and only if there exists a function $U : \mathcal{P} \times \mathcal{F} \to R$ which represents $\succsim$ and a mixture-linear and continuous function $u : \Delta(X) \to R$ and a number $\varepsilon \in [0,1]$ such that*

$$U(P, f) = \sum_{\omega \in \Omega} u(f(\omega))\, s(P)(\omega)$$

Moreover, u is unique up to positive linear transformations.

One might also wonder about generalizing Theorem 1 by dropping Invariance to Similarity Reshuffles. If that axiom is dropped, then the current proof shows that the remaining axioms imply $(*)$, where $\varphi : \mathcal{P} \to \mathcal{P}$ need not satisfy $(**)$. Instead, φ satisfies the following three conditions.

Selection: For any $P \in \mathcal{P}$, $\varphi(P) \subset P$.

Mixture linearity: For any $P, Q \in \mathcal{P}$, $\varphi(\lambda P + (1-\lambda)Q) = \lambda\varphi(P) + (1-\lambda)\varphi(Q)$.

and

Continuity: φ is continuous over $\mathcal{P}$ in the Hausdorff metric.

However, it is not clear how such a mapping is represented concretely.

5 Comparative Imprecision Aversion

The parameter ε represents the DM's *degree of imprecision aversion.*

Say that $\succsim_1$ is *more imprecision averse* than $\succsim_2$ if for any $P \in \mathcal{P}$, $p \in \Delta(\Omega)$ and $f \in \mathcal{F}$, $(\{p\}, f) \succsim_2 (P, f)$ implies $(\{p\}, f) \succsim_1 (P, f)$. This says that the two preferences coincide over probability-possibility set/act pairs with precise information , but pp set/act pair with imprecise information disliked by one is disliked by the more imprecision averse one.

Notice that this comparison is partial in the sense that the degree of imprecision aversion is comparable only between the preferences with the same ranking of risky prospects.

In the class of preferences characterized in the previous theorem, the degree of imprecision aversion is parametrized by ε.

Theorem 2 *Maintain axioms 1-10. Let* (u_1, ε_1) *and* (u_2, ε_2) *represent* $\succsim_1$ *and* $\succsim_2$*, respectively. Then,* $\succsim_1$ *is more imprecision averse than* $\succsim_2$ *if and only if there exists constants* A, B *with* $A > 0$ *such that* $u_1 = Au_2 + B$*, and* $\varepsilon_1 \geq \varepsilon_2$*.*

The degree of imprecision aversion ε can be elicited by a simple procedure. Bring two urns, each containing balls which are possibly n colors including red. The DM is told nothing about the proportions of colors in urn 1. For urn 2, she is told that the proportion of red is p. Denote a bet on red in urn i by R_i, where the bet yields payoff x^* if red is drawn from i, and $x \prec x^*$ if not. Suppose p_n is such that the DM is indifferent between R_1 and R_2 — thus p_n is called a 'precise information equivalent.' One calculates immediately that $p_n = (1 - \varepsilon)/n$, and hence $\varepsilon = 1 - np_n$. When the DM faces any other information represented by some set P, her subjective set of priors is $(1 - \varepsilon)\{s(P)\} + \varepsilon P$. Since $u(\cdot)$ can be elicited through choice between lotteries, the model permits prediction of behavior under different information.

6 Further Discussion

6.1 Second-order prior model

We compare our model and the second-order prior model described in section 1.3. In order to be precise, we introduce some technical details. Let $\mathcal{C}$ be the space of closed subsets of $\Delta(\Omega)$ and let $\mathcal{C}^* = \{P \in \mathcal{C} : \overline{intP} = P \quad \text{or} \quad |P| = 1\}$. Both are endowed with the Hausdorff metric. The second-order prior model is well-defined on $\mathcal{C}^*$.

To facilitate comparison, we take the common domain $\mathcal{P}^* \times \mathcal{F}$ where $\mathcal{P}^* = \mathcal{P} \cap \mathcal{C}^*$. Table 1 compares the two models on $P^* \times \mathcal{F}$. One critical difference is that the second-order prior model violates Independence. Another difference is that Gains via Hedging and Cardinal Invariance are violated when ϕ is nonlinear. This suggests that the imprecision aversion properties described by the two models are different. Although both models are continuous in the subdomain $\mathcal{P}^* \times \mathcal{F}$, there is an implicit difference between them also with regard to continuity. The second-order prior model cannot be continuously extended to $\mathcal{P}$ or $\mathcal{C}$, since it is not defined when the probability-possibility set has empty interior.

Axioms	Present model	Second-order prior model
Order	✓	✓
Continuity	✓	✓
Outcome Preference	✓	✓
Reduction under Precision	✓	✓
Independence	✓	×
Set Dominance	✓	✓
Act Dominance	✓	✓
Gains via Hedging	✓	✓ if ϕ linear
Cardinal Invariance	✓	✓ if ϕ linear
Invariance to SR	✓	✓ if ϕ linear and ν uniform

Table 1: Comparison of the two models on $\mathcal{P}^*$

The two models are not nested – in fact, they are disjoint. That is, there does not exist a preference order on $\mathcal{P}^*$ that can be represented both in the form described in Theorem 1 and in the form given by ($\star$). The proof is provided at the end of the appendix.

6.2 Ambiguity as sets of lotteries

An alternative way to model objective but imprecise information is to specify sets of lotteries over outcomes as the objects of choice. This approach is taken by Olszewski [13], Ahn [1] and Stinchcombe [18].

Our model reduces to the lottery set model if the DM evaluates a probability-possibility set/act pair (P, f) only via the set of distributions over outcomes (or lotteries) induced by (P, f). Formally, the condition is expressed in the form of the following axiom,

Set Reduction: For any $P, Q \in \mathcal{P}$ and $f, g \in \mathcal{F}$,

$$\{l(p, f) : p \in P\} = \{l(q, g) : q \in Q\} \Longrightarrow (P, f) \sim (Q, g).$$

When $P = Q = \{p\}$, Set Reduction asserts probabilistic sophistication in the sense of Machina-Schmeidler [12], which is satisfied in our model (see Axiom 4). However, Set Reduction is not always intuitive when information is imprecise. Two probability-possibility set/act pairs may be evaluated differently even if they induce the same sets of induced distributions. Consider the Ellsberg's urn example. There are two urns each containing 100 balls which are red, green, blue or yellow. The DM is informed that urn 1 contains only red or green balls, but knows nothing about urn 2. She holds a bet on red from the urn of her choice. She receives 100 dollars if red is drawn and nothing otherwise. If the DM cares only about the sets of induced distributions, she is indifferent between the two urns. However, she may plausibly prefer to bet on urn 1.

Appendix

A Proofs

Necessity of the axioms is routine. We show sufficiency.

A.1 Precise information case

Proposition 1 *There exists a mixture linear and continuous function $u : \Delta(X) \to R$ such that $(\{p\}, f) \succsim (\{q\}, g)$ if and only if*

$$\sum_{\omega\in\Omega} u(f(\omega))\, p(\omega) \geq \sum_{\omega\in\Omega} u(g(\omega))\, q(\omega).$$

Moreover, u is unique up to positive linear transformations.

Fix a possibility set P^*. By Outcome Preference, we can define the preference over outcome lotteries $\succsim^*$ by

$$l \succsim^* l' \qquad \text{if} \qquad (P^*, l) \succsim (P^*, l').$$

Lemma 2 *For any $l_1, l_2, l_3 \in \Delta(X)$ and $\lambda \in (0,1)$, $l_1 \succsim^* l_2$ if and only if $\lambda l_1 + (1-\lambda)l_3 \succsim^* l_2 + (1-\lambda)l_3$.*

Proof. Let $f \in \mathcal{F}$ an act defined by

$$f(\omega) = \begin{cases} l_1 & \text{if} \qquad \omega = \omega_1, \\ l_2 & \text{if} \qquad \omega = \omega_2, \\ l_3 & \text{if} \qquad \omega = \omega_3 \\ l_4 & \text{otherwise.} \end{cases}$$

and $\delta_1, \delta_2, \delta_3$ be the probability degenerate on $\omega_1, \omega_2, \omega_3$, respectively.

Then, for any $\lambda \in (0,1)$

$$\begin{aligned}
l_1 \succ^* l_2 &\Leftrightarrow l(\delta_1, f) \succ^* l(\delta_2, f) \\
&\Leftrightarrow (\{\delta_1\}, f) \succ (\{\delta_2\}, f) \\
&\Rightarrow (\{\lambda\delta_1 + (1-\lambda)\delta_3\}, f) \succ (\{\lambda\delta_2 + (1-\lambda)\delta_3\}, f) \\
&\Leftrightarrow l(\lambda\delta_1 + (1-\lambda)\delta_3, f) \succ^* l(\lambda\delta_2 + (1-\lambda)\delta_3, f) \\
&\Leftrightarrow \lambda l(\delta_1, f) + (1-\lambda) l(\delta_3, f) \succ^* \lambda l(\delta_2, f) + (1-\lambda) l(\delta_3, f) \\
&\Leftrightarrow \lambda l_1 + (1-\lambda) l_3 \succ^* \lambda l_2 + (1-\lambda) l_3
\end{aligned}$$

Thus, $\succsim^*$ satisfies the conditions for the vNM expected utility over lotteries. By Grandmont [10], it has a mixture linear and continuous representation denoted by $u : \Delta(X) \to R$, which is unique up to positive linear transformations. Without loss of generality, let $l^* \in \Delta(X)$ be such that $u(l^*) = 0$. ■

Proof of Proposition 1

Since u is mixture linear, by Reduction under Precise Information we obtain

$$\begin{aligned}
(\{p\}, f) \succsim (\{q\}, g) &\Leftrightarrow (\{p\}, l(p, f)) \succsim (\{q\}, l(q, g)) \\
&\Leftrightarrow (P^*, l(p, f)) \succsim (P^*, l(q, g)) \\
&\Leftrightarrow l(p, f) \succsim^* l(q, g) \\
&\Leftrightarrow u(l(p, f)) \geq u(l(q, g)) \\
&\Leftrightarrow \sum_{\omega \in \Omega} u(f(\omega))\, p(\omega) \geq \sum_{\omega \in \Omega} u(g(\omega))\, q(\omega).
\end{aligned}$$

This completes the proof. ■

A.2 Construction of the utility function

Proposition 2 *Given u, there exists a unique continuous function $U : \mathcal{P} \times \mathcal{F} \to R$ such that:*

1. *$(P, f) \succsim (Q, g)$ if and only if $U(P, f) \geq U(Q, f)$*
2. *$U(\{p\}, f) = \sum_{\omega \in \Omega} u(f(\omega))\, p(\omega)$ over $\Delta(\Omega) \times \mathcal{F}$*
3. *$U(\cdot, f)$ is mixture linear on $\mathcal{P}$.*

Set Dominance implies the following:

Lemma 3 *For any $P \in \mathcal{P}$ and $f \in \mathcal{F}$, $(\{\overline{p}\}, f) \succsim (P, f) \succsim (\{\underline{p}\}, f)$ for $\overline{p}, \underline{p} \in P$ such that $(\{\overline{p}\}, f) \succsim (\{p\}, f) \succsim (\{\underline{p}\}, f)$ for every $p \in P$.*

We show that under a fixed choice of act, any probability-possibility set has a *precise information equivalent*, which is a precise information equally preferable to it.

Lemma 4 *For any $P \in \mathcal{P}$ and $f \in \mathcal{F}$, there exists $p \in \Delta(\Omega)$ such that $(P, f) \sim (\{p\}, f)$.*

Proof. By the previous lemma, $(\{\overline{p}\}, f) \succsim (P, f) \succsim (\underline{p}, f)$ for $\overline{p}, \underline{p} \in P$ such that $(\{\overline{p}\}, f) \succsim (\{p\}, f) \succsim (\{\underline{p}\}, f)$ for every $p \in P$. It is immediate when $(\{\overline{p}\}, f) \sim (\{\underline{p}\}, f)$. Otherwise, Mixture Independence implies there is a unique $\lambda \in [0, 1]$ such that $(P, f) \sim (\{(1-\lambda)\overline{p} + \lambda\underline{p}\}, f)$. ■

Independence implies that equivalent precise information is preserved under mixing.

Lemma 5 *If $(P, f) \sim (\{p\}, f)$ and $(Q, f) \sim (\{q\}, f)$, then $(\lambda P + (1-\lambda)Q, f) \sim (\{\lambda p + (1-\lambda)q\}, f)$ for any $\lambda \in [0, 1]$.*

Proof of Proposition 2

For each $P \in \mathcal{P}$ and $f \in \mathcal{F}$, define $U(P, f)$ by

$$U(P, f) \equiv U(\{p\}, f)$$

where $p \in \Delta(\Omega)$ is such that $(P, f) \sim (\{p\}, f)$.

By construction, (i) this represents the preference, and (ii) $U(\{p\}, f) = \sum_{\omega \in \Omega} u(f(\omega))\, p(\omega)$ over $\Delta(\Omega) \times \mathcal{F}$. By the above lemma, (iii) $U(\cdot, f)$ is mixture-linear over $\mathcal{P}$ for each fixed f. Uniqueness follows from (ii). ■

A.3 An intermediate representation result

The aim of this subsection is to prove:

Proposition 3 *Given u and U, there exists a unique mapping $\varphi(P) : \mathcal{P} \to \mathcal{P}$ such that*

$$U(P, f) = \min_{p \in \varphi(P)} \sum_{\omega \in \Omega} u(f(\omega))\, p(\omega).$$

where φ satisfies: (i) $\varphi(P) \subset P$ for every $P \in \mathcal{P}$, (ii) continuity over $\mathcal{P}$ with respect to the Hausdorff metric, (iii) $\varphi(\lambda P + (1-\lambda)Q) = \lambda\varphi(P) + (1-\lambda)\varphi(Q)$ for any $P, Q \in \mathcal{P}$, and (iv) $\varphi(\Pi P) = \Pi\varphi(P)$ for every $P \in \mathcal{P}$ and $\Pi \in SR$.

First, we show a monotonicity condition holds.

Lemma 6 *(**Monotonicity**) For any $P \in \mathcal{P}$ and $f, g \in \mathcal{F}$,*

$$(P, f(\omega)) \succsim (P, g(\omega)) \qquad \text{for every} \quad \omega \in \Omega$$

implies $(P, f) \succsim (P, g)$.

Proof. Take $P \in \mathcal{P}$ and let $(P, f(\omega)) \succsim (P, g(\omega))$ for every $\omega \in \Omega$. Under axioms 1-5, this implies $(\{p\}, f) \succsim (\{p\}, g)$ for every $p \in P$. By Act Dominance, we get $(P, f) \succsim (P, g)$. ■

For later use, let $\succsim_f$ be the ordering on $\mathcal{P}$ conditional on f, that is induced by $\succsim$; that is,

$$P \succsim_f Q \Longleftrightarrow (P, f) \succsim (Q, f)$$

for any $P, Q \in \mathcal{P}$. Let $R_u = \{u \circ f \in R^\Omega : f \in \mathcal{F}\}$.

Lemma 7 *Given u and U, there exists a unique function $I : \mathcal{P} \times R_u \to R$ such that*

1. *$I(P, u \circ f) = U(P, f)$ for any $P \in \mathcal{P}$ and $f \in \mathcal{F}$.*
2. *$I(\cdot, x)$ is mixture linear for each $x \in R_u$.*
3. *$I(P, c\mathbf{1}) = c$ for any $c \in [\min_{l \in \Delta(X)} u(l), \max_{l \in \Delta(X)} u(l)]$.*

Proof. By Monotonicity, $I : \mathcal{P} \times R_u \to R$ is well-defined by

$$I(P, u \circ f) \equiv U(P, f).$$

In addition $I(\{p\}, u \circ f) = \sum_\omega u(f(\omega))\ p(\omega)$ for every $p \in \Delta(\Omega)$ and $f \in \mathcal{F}$. Mixture linearity follows from that of U. Monotonicity delivers $I(P, c\mathbf{1}) = c$. ■

Now we extend I to $\mathcal{P} \times R^\Omega$ by homogeneity.

Lemma 8 *For any $P \in \mathcal{P}$, $x \in R^\Omega$, $c \in R$ and $\lambda \geq 0$,*

$$I(P, \lambda x) = \lambda I(P, x).$$

Proof. First, we consider the case $0 < \lambda \leq 1$. It is immediate when x is a constant vector. Suppose x is non-constant.

Let $x = u \circ f$. Then $\lambda x = u \circ (\lambda f + (1 - \lambda) l^*)$.

By Cardinal Invariance, $\succsim_f = \succsim_{\lambda f + (1-\lambda) l^*}$. They are nondegenerate since the restricted relations over $\Delta(\Omega)$ are nondegenerate. Since both $I(\cdot, x)$ and $I(\cdot, \lambda x)$ are mixture-linear functions representing the same nontrivial order, we get cardinal equivalence

$$I(\cdot, \lambda x) = aI(\cdot, x) + b,$$

for $a > 0$ (a and b may depend on x).

When we restrict attention to singletons, then

$$\begin{aligned} I(\{p\}, x) &= px \\ I(\{p\}, \lambda x) &= \lambda px, \end{aligned}$$

which implies by varying $p$that $a = \lambda$, $b = 0$.

The result for $\lambda > 1$ and $\lambda = 0$ is immediate from the above. ■

Lemma 9 *For any $P \in \mathcal{P}$, $x \in R^{\Omega}$, $c \in R$ and $\lambda \in [0, 1]$,*

$$I(P, \lambda x + (1 - \lambda)c\mathbf{1}) = \lambda I(P, x) + (1 - \lambda)c.$$

Proof. The cases $\lambda = 0$ or 1 are immediate. Also, it is immediate when x is a constant vector. Thus, suppose $0 < \lambda < 1$ and x is a nonconstant vector. Wlog, let $x = u \circ f$ and $\lambda x + (1-\lambda)c\mathbf{1} = u \circ (\lambda f + (1-\lambda)l)$ where f is a payoff-nonconstant act. Then, f and $\lambda f + (1-\lambda)l$ are payoff nonconstant acts and induces the same and nontrivial ordering over $\mathcal{P}$.

Since both $I(\cdot, x)$ and $I(\cdot, \lambda x + (1 - \lambda)c\mathbf{1})$ are mixture-linear functions representing the same nontrivial ordering $\succsim_f = \succsim_{\lambda f+(1-\lambda)l}$, we get cardinal equivalence

$$I(\cdot, \lambda x + (1 - \lambda)c\mathbf{1}) = aI(\cdot, x) + b$$

for $a > 0$ (a and b are depending on x and c here yet).

Take any p. Then,

$$\begin{aligned} I(\{p\}, \lambda x + (1 - \lambda)c\mathbf{1}) &= \lambda px + (1 - \lambda)c \\ & \quad apx + b \end{aligned}$$

Since this is true for arbitrary p and we can take another q so that $px > qx$, we obtain $a = \lambda$ and $b = (1 - \lambda)c$. ■

Next, we show that I is concave in the second argument.

Lemma 10 *$I(P, \lambda x + (1 - \lambda)y) \geq \lambda I(P, x) + (1 - \lambda)I(P, y)$.*

Proof. It is immediate from the above lemma when either x or y is a constant vector. Suppose both are non-constant vectors.

First, we consider the case $pr_{\Delta}x \nparallel pr_{\Delta}y$, where $pr_{\Delta}x$ is the projection of x over the affine subspace of R^{Ω} spanned by $\Delta(\Omega)$.

Let $x = u \circ f$, $y = u \circ g$. Then, $\lambda x + (1 - \lambda)y = u \circ (\lambda f + (1 - \lambda)g)$.

Take any $P \in \mathcal{P}$. Let

$$\begin{aligned}\psi(P,f) &= \{p \in \Delta(\Omega) : (P,f) \sim (\{p\},f)\} \\ \psi(P,g) &= \{p \in \Delta(\Omega) : (P,g) \sim (\{p\},g)\}\end{aligned}$$

$\psi(P,f)$ is represented as the intersection of the hyperplane $\{p \in R^{\Omega} : p(u \circ f) = U(P,f)\}$ and the probability simplex $\Delta(\Omega)$.

Case 1: $\psi(P,f) \cap \psi(P,g) \neq \emptyset$. Take $p \in \psi(P,f) \cap \psi(P,g)$. Then,

$$\begin{aligned}I(P, \lambda x + (1-\lambda)y) &= U(P, \lambda f + (1-\lambda)g) \\ &\geq U(\{p\}, \lambda f + (1-\lambda)g) \\ &= \lambda U(\{p\}, f) + (1-\lambda)U(\{p\}, g) \\ &= \lambda I(P,x) + (1-\lambda)I(P,y)\end{aligned}$$

where the inequality on the second line follows from Gains via Hedging.

Case 2: Even when $\psi(P,f) \cap \psi(P,g) = \emptyset$, we can take $q \in int\Delta(\Omega)$ and sufficiently small $\eta > 0$ so that $\psi(\eta P + (1-\eta)\{q\}, f) \cap \psi(\eta P + (1-\eta)\{q\}, g) \neq \emptyset$. When $\lambda x + (1-\lambda)y$, we get $I(\eta P + (1-\eta)\{q\}, \lambda x + (1-\lambda)y) \geq \lambda I(P + (1-\eta)\{q\}, x) + (1-\lambda)I(P + (1-\eta)\{q\}, y)$. Since $I(\eta P + (1-\eta)\{q\}, z) = \eta I(P,z) + (1-\eta)qz$ for any non-constant vector z, we obtain the result. When $\lambda x + (1-\lambda)y$, Continuity delivers the desired result.

When $pr_{\Delta}x \parallel pr_{\Delta}y$, take an sequence of non-constant vectors (x^n, y^n) converging to (x,y) such that $pr_{\Delta}x^n \nparallel pr_{\Delta}y^n$ for every n. We have obtained $I(P, \lambda x^n + (1-\lambda)y^n) \geq \lambda I(P, x^n) + (1-\lambda)I(P, y^n)$ for every n. By Continuity, we obtain the desired result. ■

Thus, we can apply the Gilboa-Schmeidler [8] argument so that we obtain

Lemma 11 *For any $P \in \mathcal{P}$, there exists a unique closed (hence compact) convex set $\varphi(P)$ such that*

$$I(P,x) = \min_{p \in \varphi(P)} \sum_{\omega} x(\omega)\, p(\omega)$$

for every $x \in R^{\Omega}$.

Proof. Fix $P \in \mathcal{P}$. Then, $I(P, \cdot)$ satisfies the condition of GS. Thus, there is a closed convex set $\varphi(P) \in \mathcal{P}$ such that

$$I(P,x) = \Phi\left(\min_{p \in \varphi(P)} \sum_{\omega} x(\omega)\, p(\omega); P\right)$$

where $\Phi(\cdot; P)$ is a monotone transformation depending on P.

Recall that $I(P, c\mathbf{1}) = c$ for any c. Therefore, $\Phi(c; P) = c$ for any c, which implies $\Phi(\cdot, P)$ is an identity map, and it this true for any $P \in \mathcal{P}$. ■

Lemma 12 *For every $P \in \mathcal{P}$, $\varphi(P) \subset P$.*

Proof. Suppose $\varphi(P) \nsubseteq P$. Then, there exists $f \in \mathcal{F}$ such that

$$\begin{aligned} U(P, f) &= \min_{p \in \varphi(P)} \sum_{\omega} u(f(\omega))\, p(\omega) \\ &< \min_{p \in P} \sum_{\omega} u(f(\omega))\, p(\omega) \\ &= U(\{\underline{p}_f\}, f), \end{aligned}$$

which contradicts Lemma 3. ■

Thus, the mapping $\varphi : \mathcal{P} \to \mathcal{P}$ with $\varphi(P) \subset P$ is well-defined.

Lemma 13 *For any $P, P' \in \mathcal{P}$ and $\lambda \in [0, 1]$, $\varphi(\lambda P + (1 - \lambda)P') = \lambda\varphi(P) + (1 - \lambda)\varphi(P')$.*

Proof. By construction,

$$U(\lambda P + (1 - \lambda)P', f) = \min_{p \in \varphi(\lambda P + (1-\lambda)P')} \sum_{\omega} u(f(\omega))\, p(\omega).$$

for any $f \in \mathcal{F}$.

By mixture-linearity of U over $\mathcal{P}$,

$$\begin{aligned} \lambda U(P, f) + (1 - \lambda)U(P', f) &= \lambda \min_{p \in \varphi(P)} \sum_{\omega} u(f(\omega))\, p(\omega) + (1 - \lambda) \min_{p \in \varphi(P')} \sum_{\omega} u(f(\omega))\, p(\omega) \\ &= \min_{p \in \lambda\varphi(P) + (1-\lambda)\varphi(P')} \sum_{\omega} u(f(\omega))\, p(\omega) \end{aligned}$$

for any $f \in \mathcal{F}$. By uniqueness of $\varphi(\cdot)$, we obtain the result. ■

Lemma 14 *The mapping $\varphi : \mathcal{P} \to \mathcal{P}$ is continuous with respect to the Hausdorff metric.*

Proof. Let $\{P^n\}$ be a sequence in $\mathcal{P}$ converging to $P \in \mathcal{P}$. Because $\mathcal{P}$ is compact, it is wlog to assume that $\{\varphi(P^n)\}$ is convergent. Suppose $\varphi^* \equiv \lim_{n\to\infty} \varphi(P^n) \neq \varphi(P)$. Then there exists $f \in \mathcal{F}$ such that

$$\begin{aligned} U(P, f) &= \min_{p \in \varphi(P)} \sum_{\omega} u(f(\omega))\, p(\omega) \\ &> \min_{p \in \varphi^*} \sum_{\omega} u(f(\omega))\, p(\omega) \\ &= \lim_{n\to\infty} \min_{p \in \varphi(P^n)} \sum_{\omega} u(f(\omega))\, p(\omega) \\ &= \lim_{n\to\infty} U(P^n, f), \end{aligned}$$

which is a contradiction to Continuity. ■

Lemma 15 *Bistochastic matrix* Π *is a similarity reshuffle if and only if there exists* $\lambda \in (0,1]$ *such that* $\Pi^t\Pi = \lambda I + \frac{1-\lambda}{|\Omega|}E$*, where* I *is the identity matrix and* E *is a matrix in which all the entries are 1.*

Proof. Recall $e = (\frac{1}{|\Omega|}, \cdots, \frac{1}{|\Omega|})$.
($\Longrightarrow$): Take any $p, q \in \Delta(\Omega)$. Then,

$$\begin{aligned}\langle \Pi p, \Pi q\rangle &= 2\left\langle \frac{\Pi p+\Pi q}{2} - e, \frac{\Pi p + \Pi q}{2} - e\right\rangle - \frac{1}{2}\langle \Pi p - e, \Pi p - e\rangle - \frac{1}{2}\langle \Pi q - e, \Pi q - e\rangle \\ &\quad + \langle \Pi p, e\rangle + \langle \Pi q, e\rangle - \langle e, e\rangle\end{aligned}$$

Since $\Pi e = e$ and $\langle \Pi p, e\rangle = \langle p, e\rangle = \frac{1}{|\Omega|}$ for any p, by assumption the right-hand-side of the above equality becomes

$$\begin{aligned}&\lambda\left[2\left\langle \frac{p+q}{2} - e, \frac{p+q}{2} - e\right\rangle - \frac{1}{2}\langle p-e, p-e\rangle - \frac{1}{2}\langle q-e, q-e\rangle + \langle p, e\rangle + \langle q, e\rangle - \langle e, e\rangle\right] \\ &+(1-\lambda)\left[\langle p, e\rangle + \langle q, e\rangle - \langle e, e\rangle\right] \\ =\ &\lambda\langle p, q\rangle + (1-\lambda)/|\Omega|\end{aligned}$$

Thus, we obtain

$$\begin{aligned}\langle \Pi p, \Pi q\rangle &= p^t\Pi^t\Pi q \\ &= \lambda\langle p, q\rangle + (1-\lambda)/|\Omega|\end{aligned}$$

for any $p, q \in \Delta(\Omega)$. Pick $p = \delta_\omega$, $q = \delta_{\omega'}$ where δ_ω and $\delta_{\omega'}$ are probabilities degenerated on ω, ω', respectively. Then, for all the column vectors $\Pi_1, \cdots, \Pi_{|\Omega|}$, we obtain

$$\begin{aligned}\Pi^t_\omega\Pi_{\omega'} &= \lambda \qquad \text{when} \quad \omega \neq \omega' \\ \Pi^t_\omega\Pi_\omega &= \lambda + (1-\lambda)/|\Omega|.\end{aligned}$$

($\Longleftarrow$): By the converse argument of the above. ■

Lemma 16 *For any* $P \in \mathcal{P}$ *and* $\Pi \in SR$*,* $\varphi(\Pi P) = \Pi\varphi(P)$*.*

Proof. Suppose $\varphi(\Pi P) \nsubseteq \Pi\varphi(P)$. Then, there is $x \in R^\Omega$ such that

$$\min_{p\in\Pi\varphi(P)}\sum_\omega y(\omega)\ p(\omega) > \min_{p'\in\varphi(\Pi P)}\sum_\omega y(\omega)\ p'(\omega)$$

By taking $y = \Pi x$, both sides are written as

$$\min_{p\in\varphi(P)}\sum_\omega (\Pi x)(\omega)\ (\Pi p)(\omega) > \min_{p'\in\varphi(\Pi P)}\sum_\omega (\Pi x)(\omega)\ p'(\omega) \qquad (*)$$

By homogeneity with respect to x, wlog we can set $x = u \circ f$ and for some $f \in \mathcal{F}$. Take $p^* \in \arg\min_{p \in \varphi(P)} \sum_\omega x(\omega)\ p(\omega)$. Since

$$\sum_\omega (\Pi x)(\omega)\ (\Pi p)(\omega) = \lambda \sum_\omega x(\omega)\ p(\omega) + \frac{1-\lambda}{|\Omega|} \sum_\omega x(\omega),$$

we have $p^* \in \arg\min_{p \in \varphi(P)} \sum_\omega (\Pi x)(\omega)\ (\Pi p)(\omega)$. Thus, the left hand side of $(*)$ is equal to $U(\{\Pi p^*\}, \Pi f)$. On the other hand, the right hand side of $(*)$ is $U(\Pi P, \Pi f)$. Thus, $U(\{\Pi p^*\}, \Pi f) > U(\Pi P, \Pi f)$

By definition of p^*, we have $U(\{p^*\}, f) = U(P, f)$. This contradicts to the Invariance to Similarity Reshuffles.

We similarly obtain a contradiction for the case $\varphi(\Pi P) \not\supseteq \Pi\varphi(P)$. ■

A.4 Analysis of the prior selection mapping

Let $\mathbf{1} = (1, \cdots, 1)$ and $e = \frac{1}{|\Omega|}\mathbf{1}$. For later use, we show the lemma below.

Lemma 17 *Let $F : \Delta(\Omega) \to \Delta(\Omega)$ be a mixture-linear mapping satisfying $F(e) = e$. Then there is a unique doubly stochastic matrix Π such that $F(p) = \Pi p$ for every $p \in \Delta(\Omega)$.*

Proof. Given such F, define Π by $\Pi_{ij} = F_i(\delta_j)$ where δ_j is a probability which assigns unit mass on state $j \in \Omega$. By mixture linearity, Π represents F.

Suppose there are two matrices Π and Π' which represent F. If $\Pi_{ij} \neq \Pi'_{ij}$ for some $i, j \in \Omega$, this leads to $F_i(\delta_j) = \Pi_{ij} \neq \Pi'_{ij} = F_i(\delta_j)$, a contradiction. Thus Π is unique.

If $\Pi_{ij} < 0$ for some $i, j \in \Omega$, this leads to $F_i(\delta_j) < 0$, which is a contradiction.

For any $j \in \Omega$, $\Pi\delta_j = (\Pi_{ij})_{i\in\Omega} \in \Delta(\Omega)$. Therefore, $\sum_{i\in\Omega} \Pi_{ij} = 1$ for each $j \in \Omega$.

Since $\Pi e = e$, for each $i \in N$, $\frac{1}{|\Omega|}\sum_{j\in\Omega} \Pi_{ij} = \frac{1}{|\Omega|}$. Therefore, $\sum_{j\in\Omega} \Pi_{ij} = 1$ for each $i \in \Omega$. ■

Now define $\varphi^* : \mathcal{P} - \{e\} \to \mathcal{P} - \{e\}$ by

$$\varphi^*(K) = \varphi(K + \{e\}) - \{e\}$$

Lemma 18 *For any $K \in \mathcal{P} - \{e\}$ and $\lambda \geq 0$ with $\lambda K \in \mathcal{P} - \{e\}$, $\varphi^*(\lambda K) = \lambda\varphi^*(K)$.*

Proof. The case of $\lambda = 0$ or 1 is obvious. Let $\lambda \in (0, 1)$. Then,

$$\begin{aligned}
\varphi^*(\lambda K) &= \varphi(\lambda K + \{e\}) - \{e\} \\
&= \varphi(\lambda(K + \{e\}) + (1-\lambda)\{e\}) - \{e\} \\
&= \lambda\varphi(K + \{e\}) + (1-\lambda)\varphi(\{e\}) - \{e\} \\
&= \lambda\varphi(K + \{e\}) + (1-\lambda)\{e\} - \{e\} \\
&= \lambda(\varphi(K + \{e\}) - \{e\}) \\
&= \lambda\varphi^*(K).
\end{aligned}$$

The case of $\lambda > 1$ is immediate from the above. ■

Let H_e be the $|\Omega| - 1$ dimensional linear subspace of R^{Ω} which is orthogonal to e. Let $\mathcal{K}_e$ be the family of compact convex subsets of H_e, endowed with the Hausdorff metric. By the above lemma, we extend φ^* to $\mathcal{K}_e$. This preserves continuity in the Hausdorff metric.

Lemma 19 *For any $K, K' \in \mathcal{K}_e$, $\varphi^*(K + K') = \varphi^*(K) + \varphi^*(K')$. In particular, $\varphi^*(K + \{z\}) = \varphi^*(K) + \{z\}$.*

Proof. Take sufficiently small $\lambda > 0$, then $\lambda K, \lambda K' \in \mathcal{P} - \{e\}$. By homogeneity,

$$\varphi^*(K + K') = \frac{2}{\lambda}\varphi^*\left(\frac{\lambda K + \lambda K'}{2}\right)$$

Then, we have

$$\begin{aligned}
\varphi^*\left(\frac{\lambda K + \lambda K'}{2}\right) &= \varphi\left(\frac{\lambda K + \{e\}}{2} + \frac{\lambda K' + \{e\}}{2}\right) - \{e\} \\
&= \frac{1}{2}\varphi(\lambda K + \{e\}) + \frac{1}{2}\varphi(\lambda K' + \{e\}) - \{e\} \\
&= \frac{\varphi(\lambda K + \{e\}) - \{e\}}{2} + \frac{\varphi(\lambda K' + \{e\}) - \{e\}}{2} \\
&= \frac{1}{2}\varphi^*(\lambda K) + \frac{1}{2}\varphi^*(\lambda K') \\
&= \frac{\lambda}{2}\varphi^*(K) + \frac{\lambda}{2}\varphi^*(K'),
\end{aligned}$$

which gives the result. ■

For each $\Pi \in SR$, define a mapping $T_\Pi : \Delta(\Omega) - \{e\} \to \Delta(\Omega) - \{e\}$ by

$$T_\Pi(x) = \Pi(x + e) - e.$$

Since $x + e \in \Delta(\Omega)$ implies $\Pi(x + e)$, the mapping indeed satisfies $T_\Pi(x) \in \Delta(\Omega) - \{e\}$. By nature of Π, we have (i) $T_\Pi(\lambda x) = \lambda T_\Pi(x)$ for any λ with $\lambda x \in \Delta(\Omega) - \{e\}$, (ii) $T_\Pi(x+y) = T_\Pi(x) + T_\Pi(y)$, and (iii) there exists $\lambda_\Pi \in (0, 1)$ such that $\|T_\Pi(x)\| = \lambda_\Pi \|x\|$ for every $x \in \Delta(\Omega) - \{e\}$. By (i), we can extend T_Π to the whole linear subspace H_e.

We say that a linear transformation T is a *sub-similarity* if $T(\Delta(\Omega) - \{e\}) \subset \Delta(\Omega) - \{e\}$ and there exists $\lambda_T \in (0, 1]$ such that $\|T(x)\| = \lambda_T \|x\|$ for any $x \in H_e$.

Conversely to the above, any sub-similarity $T : H_e \to H_e$ has a corresponding SR. For T, define $F_T : \Delta(\Omega) \to \Delta(\Omega)$ by

$$F_T(p) = T(p - e) + e.$$

Then, it is easy to see that F_T takes values in $\Delta(\Omega)$ and is mixture linear and $F_T(e) = e$. By the previous lemma, it has a representation by a doubly stochastic matrix Π_T and $F_T(p) = \Pi(p)$. Since F_T satisfies (iii), Π_T is an SR.

Now we show that φ^* is equivariant in sub-similarities.

Lemma 20 *For any sub-similarity $T : H_e \to H_e$, $\varphi^*(TK) = T\varphi(K)$.*

Proof. By homogeneity of φ^*, wlog we can take $K \in \mathcal{P} - \{e\}$. By the above lemma T has a corresponding SR Π_T and $T(x) = \Pi_T(x+e) - e$ for any $x \in \Delta(\Omega) - \{e\}$.

Then,

$$\begin{aligned}
\varphi^*(T(K)) &= \varphi(T(K) + \{e\}) - \{e\} \\
&= \varphi(\Pi_T(K + \{e\}) - \{e\} + \{e\}) - \{e\} \\
&= \varphi(\Pi_T(K + \{e\})) - \{e\} \\
&= \Pi_T\varphi(K + \{e\}) - \{e\} \\
&= T(\varphi(K)).
\end{aligned}$$

■

A linear transformation $I : H_e \to H_e$ is called *isometry* if $\|I(x)\| = \|x\|$. Let $\mathcal{I}$ be the set of isometries. For any isometry $I \in \mathcal{I}$, its positive scaler multiplication λI where $\lambda > 0$ is chosen so that $\lambda I(\Delta(\Omega) - \{e\}) \subset \Delta(\Omega) - \{e\}$ is a sub-similarity. Conversely, any isometry is obtained from a sub-similarity by reversing the above procedure.

By homogeneity of φ^*, we obtain

Lemma 21 *The mapping φ^* is equivariant in isometries. That is, for any isometry $I \in \mathcal{I}$, $\varphi^*(I(K)) = I(\varphi^*(K))$.*

The $|\Omega| - 1$ dimensional Euclidian space $R^{|\Omega|-1}$ is the image of the linear subspace H_e by some isometry. Let $J : H_e \to R^{|\Omega|-1}$ be the isometry. All the relevant operations are preserved under isometry. Let $\mathcal{K}^{|\Omega|-1}$ be the space of compact convex subsets of $R^{|\Omega|-1}$. The space $\mathcal{K}^{|\Omega|-1}$ is also the image of $\mathcal{K}_e$ by the isometry. Define $\varphi^{**} : \mathcal{K}^{|\Omega|-1} \to \mathcal{K}^{|\Omega|-1}$ by

$$\varphi^{**}(K) = J(\varphi^*(J^{-1}(K))).$$

Then, φ^{**} is continuous, additive and equivariant in isometries in $R^{|\Omega|-1}$ and translations, and satisfies $\varphi^{**}(K) \subset K$ for any $K \in \mathcal{K}^{|\Omega|-1}$.

Let $W = \{w \in R^{|\Omega|-1} : \|w\| = 1\}$ be the $|\Omega| - 2$ dimensional unit sphere. For a compact convex set $K \in \mathcal{K}^{|\Omega|-1}$, its *Steiner point* is defined by

$$s^{**}(K) = \int_W w h_K(w)\, \mu(dw)$$

where h_K is the support function given by $h_K(v) = \max_{z \in K} zw$ and μ is normalized rotation invariant measure over W.

Case 1 $|\Omega| = 3$: Since image of a segment is its subsegment, we can apply Theorem 1.8 (b) in Schneider [16] so that we obtain

$$\varphi^{**}(K) = \varepsilon T_1 \left[K - s^{**}(K)\right] + \delta T_2 \left[-K + s^{**}(K)\right] + \{s^{**}(K)\}$$

with $\varepsilon \geq 0$, $\delta \geq 0$ and T_1, T_2 being some two dimensional rotation matrices.

Consider a segment with midpoint 0. Since its image is its subsegment, it must be the case that $(T_1, T_2) = (1, 1)$ or $(1, -1)$ or $(-1, 1)$ or $(-1, -1)$. Thus, wlog

$$\varphi^{**}(K) = \varepsilon \left[K - s^{**}(K)\right] + \delta \left[-K + s^{**}(K)\right] + \{s^{**}(K)\}$$

We show $\varepsilon \in [0, 1]$ and $\delta = 0$ below together with the next case.

Case 2 $|\Omega| \geq 4$: Since $\varphi^{**}(K) \subset K$ for any $K \in \mathcal{K}^{|\Omega|-1}$, the image of any segment is its subsegment. Thus we can apply Theorem 1.8 (b) in Schneider [16] so that we obtain

$$\varphi^{**}(K) = \varepsilon \left[K - s^{**}(K)\right] + \delta \left[-K + s^{**}(K)\right] + \{s^{**}(K)\}$$

with $\varepsilon \geq 0$, $\delta \geq 0$.

We show $\varepsilon \in [0, 1]$ and $\delta = 0$. Since $\varphi^{*}(K) \subset K$ for any K, ε cannot exceed 1. Now consider a family of triangles

$$K_\theta = \{(x_1, x_2, 0, \cdots, 0) \in R^{|\Omega|-1} : x_2 \leq \frac{\cos\theta}{\sin\theta} x_1,\ x_2 \geq -\frac{\cos\theta}{\sin\theta} x_1,\ x_1 \leq \sin\theta\}$$

indexed by $0 < \theta < \frac{\pi}{2}$. Then we have $s^{**}(K_\theta) = (\frac{\pi-\theta}{\pi} \sin\theta, 0, 0, \cdots, 0)$. Let $\overline{x}_1(K_\theta) = \max_{x \in \varphi^{**}(K_\theta)} x_1$. We get $\overline{x}_1(K_\theta) = \frac{\pi-\theta}{\pi} \sin\theta + \varepsilon \frac{\theta}{\pi} \sin\theta + \delta \frac{\pi-\theta}{\pi} \sin\theta$. Since $\varphi^{**}(K_\theta) \subset K_\theta$, this cannot exceed $\sin\theta$. Since $\sin\theta$ is positive, we can divide both sides of $\overline{x}_1(K_\theta) \leq \sin\theta$ by $\sin\theta$ and by arranging we get

$$\delta \leq \frac{\frac{\theta}{\pi}}{1 - \frac{\theta}{\pi}} (1 - \varepsilon).$$

Since this is true for any $\theta \in (0, \frac{\pi}{2})$, we obtain $\delta = 0$.

Thus

$$\varphi^{**}(K) = \varepsilon\,[K - s^{**}(K)] + \{s^{**}(K)\}$$

with $\varepsilon \in [0,1]$. Since Steiner point and every relevant operation are preserved by isometry, we obtain

$$\varphi(P) = \varepsilon\,[P - s(P)] + \{s(P)\}.$$

A.5 Comparative imprecision aversion

Below all preferences are assumed to satisfy the conditions of Theorem 1.

Lemma 22 *Let $\succsim_1$ be more imprecision averse than $\succsim_2$. Then $(\{p\}, f) \succsim_2 (\{p'\}, f)$ if and only if $(\{p\}, f) \succsim_1 (\{p'\}, f)$ for every $f \in \mathcal{F}$.*

Proof. 'Only if' part is obvious by applying the definition to the singleton case. We prove the other direction. Suppose $(\{p\}, f) \succ_1 (\{p'\}, f)$. This satisfies the hypothesis of the contrapositive of definition applied to the singleton case (Take $P = \{p\}$ and negate $(\{p'\}, f) \succsim_1 (P, f)$). Thus, $(\{p\}, f) \succ_2 (\{p'\}, f)$. This delivers the desired result by Continuity. ∎

Lemma 23 *Assume continuity. Let $\succsim_1$ be more imprecision averse than $\succsim_2$. Then, they have the same preference over outcome, i.e., $\succsim_1^* = \succsim_2^*$.*

Proof. Take any $l, l' \in \Delta(X)$. Let A be a nonempty nonuniversal subset of Ω. Define $f \in \mathcal{F}$ by

$$f(\omega) = \begin{cases} l & \text{if} \quad \omega \in A, \\ l' & \text{otherwise.} \end{cases}$$

Let $l \succsim_1^* l'$. This is equivalent that $(\delta_A, f) \succsim_1 (\delta_{\Omega \setminus A}, f)$. By the above lemma, this holds iff $(\delta_A, f) \succsim_2 (\delta_{\Omega \setminus A}, f)$, which holds iff $l \succsim_2^* l'$. ∎

Let u_1 and u_2 represent $\succsim_1^*$ and $\succsim_2^*$, respectively. By the above lemma, they are cardinally equivalent. Wlog let $u_1 = u_2 = u$ below.

Proof of Theorem 2

Proof. ('$\Longrightarrow$') Take $P \in \mathcal{P}$ and $f \in \mathcal{F}$. Choose $q \in \Delta(\Omega)$ so that

$$(1 - \varepsilon_2) \sum_{\omega \in \Omega} u(f(\omega))\; s(P)(\omega) + \varepsilon_2 \min_{p \in P} \sum_{\omega \in \Omega} u(f(\omega))\; p(\omega) = \sum_{\omega \in \Omega} u(f(\omega))\; q(\omega)$$

By comparative imprecision aversion,

$$(1 - \varepsilon_1) \sum_{\omega \in \Omega} u(f(\omega))\; s(P)(\omega) + \varepsilon_1 \min_{p \in P} \sum_{\omega \in \Omega} u(f(\omega))\; p(\omega) \leq \sum_{\omega \in \Omega} u(f(\omega))\; q(\omega)$$

Combining the above two formulas delivers

$$(\varepsilon_1 - \varepsilon_2)\left(\sum_{\omega\in\Omega} u(f(\omega))\, s(P)(\omega) - \min_{p\in P}\sum_{\omega\in\Omega} u(f(\omega))\, p(\omega)\right) \geq 0$$

Since $\sum_{\omega\in\Omega} u(f(\omega))\, s(P)(\omega) \geq \min_{p\in P}\sum_{\omega\in\Omega} u(f(\omega))\, p(\omega)$ and it is strict for some P and f, we obtain $\varepsilon_1 \geq \varepsilon_2$.

('$\Longleftarrow$') Let $\varepsilon_1 \geq \varepsilon_2$. Take $P \in \mathcal{P}$, $q \in \Delta(\Omega)$ and $f \in \mathcal{F}$. Let

$$(1-\varepsilon_2)\sum_{\omega\in\Omega} u(f(\omega))\, s(P)(\omega) + \varepsilon_2 \min_{p\in P}\sum_{\omega\in\Omega} u(f(\omega))\, p(\omega) \leq \sum_{\omega\in\Omega} u(f(\omega))\, q(\omega)$$

From $\varepsilon_1 \geq \varepsilon_2$ and $\sum_{\omega\in\Omega} u(f(\omega))\, s(P)(\omega) \geq \min_{p\in P}\sum_{\omega\in\Omega} u(f(\omega))\, p(\omega)$, we obtain

$$(1-\varepsilon_1)\sum_{\omega\in\Omega} u(f(\omega))\, s(P)(\omega) + \varepsilon_1 \min_{p\in P}\sum_{\omega\in\Omega} u(f(\omega))\, p(\omega) \leq \sum_{\omega\in\Omega} u(f(\omega))\, q(\omega)$$

■

A.6 Comparison with the second-order prior model

The second-order prior model satisfies the following condition on $\mathcal{P}^* \times \mathcal{F}$:

Convex Disjoint Union Betweeness: For any non-singleton $P, Q \in \mathcal{P}^*$ with $P \cup Q \in \mathcal{P}^*$ and $int(P \cap Q) = \emptyset$ and $f \in \mathcal{F}$,

$$(P, f) \succsim (Q, f) \Longrightarrow (P, f) \succsim (P \cup Q, f) \succsim (Q, f).$$

The following is a stronger condition.

Lower Union Indifference: For any non-singleton $P, Q \in \mathcal{P}^*$ with $P \cup Q \in \mathcal{P}^*$ and $int(P \cap Q) = \emptyset$ and $f \in \mathcal{F}$,

$$(P, f) \succsim (Q, f) \Longrightarrow (P \cup Q, f) \sim (Q, f)$$

Lemma 24 *The second-order prior model satisfies Convex Disjoint Union Betweeness on $\mathcal{P}^* \times \mathcal{F}$.*

Proof. Let $\mu(P; f) = \int_P \phi(E_p[u \circ f])\, \nu(dp)$ for $P \in \mathcal{C}$. Then $\mu(\,\cdot\,; f)$ is an additive set function over $\mathcal{C}$ for each $f \in \mathcal{F}$ such that $int(P) = \emptyset$ implies $\mu(P, f) = 0$.

Take non-singleton $P, Q \in \mathcal{P}^*$ with $P \cup Q \in \mathcal{P}^*$ and $int(P \cap Q) = \emptyset$, and suppose that $(P, f) \succsim (Q, f)$. By assumption,

$$\frac{\mu(P, f)}{\nu(P)} \geq \frac{\mu(Q, f)}{\nu(Q)}.$$

Now consider $P \cup Q$. Then, simple calculation delivers

$$\frac{\mu(P,f)}{\nu(P)} \geq \frac{\mu(P \cup Q, f)}{\nu(P \cup Q)} = \frac{\mu(P,f) + \mu(Q,f)}{\nu(P) + \nu(Q)} \geq \frac{\mu(Q,f)}{\nu(Q)}$$

■

Next, we show that our model 'almost always' violates Convex Disjoint Union Betweeness.

Lemma 25 *The representation given in Theorem 1 satisfies Convex Disjoint Union Betweeness on $\mathcal{P}^* \times \mathcal{F}$ only if $\varepsilon = 1$.*

Proof. When $\varepsilon = 1$, the representation obviously satisfies Lower Union Indifference, which is stronger than Convex Disjoint Union Betweeness.

Now assume $\varepsilon < 1$. Define $f \in \mathcal{F}$ by

$$f(\omega) = \begin{cases} l & \text{if} \quad \omega = \omega_1, \\ l^* & \text{otherwise.} \end{cases}$$

Wlog, take $u(l) > u(l^*) = 0$. Let $\delta_1, \delta_2, \delta_3$ be the probability measures degenerate on $\omega_1, \omega_2, \omega_3$, respectively. Disregard the other states wlog. Take $P = con\{\delta_1, \delta_2, \frac{\delta_2+\delta_3}{2}\}$ and $Q = con\{\delta_1, \delta_3, \frac{\delta_2+\delta_3}{2}\}$. Then, $s(P) = (\frac{10}{24}, \frac{11}{24}, \frac{3}{24})$ and $\varphi(P) = \{(p_1, p_2, p_2) : p_1 \geq \frac{10(1-\varepsilon)}{24}, p_2 - p_3 \geq \frac{8(1-\varepsilon)}{24}, p_3 \geq \frac{3(1\varepsilon)}{24}\}$. Similarly, $s(Q) = (\frac{10}{24}, \frac{3}{24}, \frac{11}{24})$ and $\varphi(Q) = \{(p_1, p_2, p_2) : p_1 \geq \frac{10(1-\varepsilon)}{24}, p_2 \geq \frac{3(1\varepsilon)}{24}, p_3 - p_2 \geq \frac{8(1-\varepsilon)}{24}\}$. Thus, $U(P,f) = U(Q,f) = \frac{10(1-\varepsilon)}{24}u(l)$.

On the other hand, since $P \cup Q = con\{\delta_1, \delta_2, \delta_3\}$, we have $s(P \cup Q) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and $\varphi(P \cup Q) = \{p_1, p_2, p_2) : p_1, p_2, p_3 \geq \frac{1-\varepsilon}{3}\}$. Thus $U(P \cup Q, f) = \frac{1-\varepsilon}{3}u(l) < \frac{10(1-\varepsilon)}{24}u(l)$, which violates Convex Disjoint Union Betweeness. Also it violates Lower Union Indifference. ■

The only remaining possibility of intersection between the two models is in the class of preferences satisfying Lower Union Indifference. However, the second-order prior model excludes Lower Union Indifference.

Lemma 26 *The second-order prior model violates Lower Union Indifference.*

Proof. Define $f \in \mathcal{F}$ by

$$f(\omega) = \begin{cases} l & \text{if} \quad \omega = \omega_1, \\ l^* & \text{otherwise.} \end{cases}$$

Wlog take $u(l) > u(l^*) = 0$.

Take $P = \{p \in \Delta(\Omega) : \langle u \circ f, p \rangle \geq \frac{1}{2}u(l)\}$ and $Q = \{p \in \Delta(\Omega) : \langle u \circ f, p \rangle \leq \frac{1}{2}u(l)\}$. Then $U(P,f) > U(P \cup Q, f) > U(Q,f)$, which violates Lower Union Indifference. ■

References

[1] Ahn, D.S., Ambiguity without a state space, Working paper, Stanford University, 2003.

[2] Anscombe, F., and R.J. Aumann, A Definition of Subjective Probability, *Annals of Mathematical Statistics* 34 (1963) 199-205.

[3] Birkoff, G., Three Observations on Linear Algebra, *Univ. Nat. Tucuman Ser.* A5 (1946) 147-151.

[4] Bolker, E.D., Functions Resembling Quotients of Measures, *Transactions of the American Mathematical Society* 124 (1966) 292-312.

[5] Castagnoli, E., F. Maccheroni, and M. Marinacci, Insurance Premia Consistent with the Market, mimeo, ICER, 2002.

[6] Damiano, E., Choice under Limited Uncertainty, mimeo, Yale University, 1999.

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

[8] Gilboa, I., D. Schmeidler, Maxmin Expected Utility with Non-unique Priors, *Journal of Mathmatical Economics* 18 (1989) 141-153.

[9] Gajdos, T., J.-M. Tallon, and J.-C. Vergnaud, Decision Making with Imprecise Probabilistic Information, 2002, forthcoming in *Journal of Mathematical Economics*.

[10] Grandmont, J.M.: Continuity Properties of a von Neumann-Morgenstern Utility, *Journal of Economic Theory* 4 (1972), 45-57.

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

[12] Machina, M., D. Schmeidler, A More Robust Definition of Subjective Probability, *Econometrica*, 60 (1992), 745-780.

[13] Olszewski, W., Preferences over Sets of Lotteries, Working paper, Northwestern University, 2002.

[14] Savage, L., The Foundations of Statistics, New York: Wiley, 1954.

[15] Schmeidler, D., Subjective Probability and Expected Utility without Additivity, *Econometrica*, 57 (1989), 571-587.

[16] Schneider, R., Equivariant Endomorphisms of the Space of Convex Bodies, *Transactions of the American Mathmatical Society* 194 (1974) 53-78.

[17] Schneider, R., Convex Bodies: the Brunn-Minkowski Theory, Cambridge, Cambridge University Press, 1993.

[18] Stinchcombe, M.B., Choice and Games with Ambiguity as Sets of Probabilities, Working paper, University of Texas, Austin, 2003.

[19] von Neumann, J. and Morgenstern, O., Theory of Games and Economic Behavior, Princeton University Press, Princeton, 1944.

[20] Wang, T., A Class of Multiple-prior Preferences, Working paper, University of British Columbia, 2003.