Attitude toward imprecise information [*]

Thibault Gajdos, Takashi Hayashi, Jean-Marc Tallon,

Jean-Christophe Vergnaud

February 2007

Abstract

This paper presents an axiomatic model of decision making under uncertainty which incorporates objective but imprecise information. 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 decision maker is told that the true probability law lies in P and is assumed to rank pairs of the form (P, f) where f is an act mapping states into outcomes. The representation result delivers maxmin expected utility at each probability-possibility set. The model explains how "beliefs" vary with information: there is a mapping that gives for each probability-possibility set the revealed set of probability distributions. This allows both expected utility when the set is reduced to a singleton and extreme pessimism when the decision maker takes the worst case scenario in the entire probability-possibility set. We define a notion of comparative imprecision aversion and show it is characterized by inclusion of the sets of revealed probability distributions, irrespective of the utility functions that capture risk attitude. We also identify an explicit attitude toward imprecision that underlies usual hedging axioms. Finally, we characterize, under extra axioms, a more precise functional form, in which the set of revealed probability distributions 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 preferences.

Keywords: Imprecise information, imprecision aversion, multiple priors, Steiner point.

JEL Number: D81.

1 Introduction

In many problems of choice under uncertainty, some information is available to the decision maker. Yet, this information is often far from being sufficiently precise to allow the decision maker to come up with an estimate of a probability distribution over the relevant states of nature. The archetypical example of such a situation is the so-called Ellsberg paradox (Ellsberg (1961)), in which subjects are given some imprecise information concerning the composition of an urn and are then asked to choose among various bets on the color of a ball drawn from that urn. The usual findings in this experiment is that most subjects' choices can not be rationalized if one assumes they hold probabilistic beliefs. Rather, one should assume that subjects have "multiple beliefs". In a very influential paper, Gilboa and Schmeidler (1989) axiomatized the following decision criterion, that allows to "solve" Ellsberg paradox: an act f is preferred to an act g if and only if there exists a set of priors $\mathcal{C}$ and a utility function u such that

$$\min_{p \in \mathcal{C}} \int u \circ f dp \geq \min_{p \in \mathcal{C}} \int u \circ g dp.$$

In this functional form, $\mathcal{C}$ is usually interpreted as the decision maker's set of beliefs. To go back to Ellsberg experiment, if one takes as the set of beliefs the distributions of the balls compatible with the information given to the subjects, then the functional with that set captures *via* the min operator an extreme form of pessimism.

In this paper, we look for ways to explain how revealed beliefs (in the sense of Gilboa and Schmeidler) are related to objective information which, in our analysis as in Ellsberg experiments, comes as a set of probability distributions. To that end, we present an axiomatic model of decision making which incorporates objective but imprecise information as a variable. On top of linking "revealed beliefs" to objective information, it also allows to uncover which type of information is valued by the decision maker.

Preferences bear on couples (P, f) where P is a set of probability distributions over the

state space (hereafter *probability-possibility sets*) and f is an act. This means that, at least conceptually, we allow decision makers to compare the same act in different informational settings. The motivation for this formalization can be best understood going back to Ellsberg's two urns example. In urn 1 there is a known proportion of black and white balls (50-50) while in urn 2, the composition is unknown. The decision maker has the choice to bet on black in urn 1 or on black in urn 2. The action (bet on black) itself is the same in the two cases but the information has changed from a given probability distribution $(\frac{1}{2}, \frac{1}{2})$ (urn 1) to the simplex (urn 2).

In an Anscombe and Aumann (1963) framework, we axiomatize the following functional form: for two probability-possibility sets P and Q and two acts f and g, (P, f) is preferred to (Q, g) if and only if

$$\min_{p \in \varphi(P)} \int u \circ f \, dp \geq \min_{p \in \varphi(Q)} \int u \circ g \, dp.$$

For a fixed probability-possibility set, this is identical to maxmin expected utility type à la Gilboa and Schmeidler (1989): in their setting, the (un-modeled) prior information that the decision maker has is given. The novel aspect here is that the link between the information possessed by the decision maker and the set of revealed probabilities is made explicit. Properties of the mapping φ, e.g. $\varphi(P) \subset P$, can be related to some further natural axioms, expressing for instance the idea that the information is considered by the decision maker to be reliable. Now, φ also embodies the decision maker's attitude toward imprecision of the information. Hence, the interpretation of $\varphi(P)$ as a set of beliefs is not warranted; we will rather call it in a more neutral way, the set of revealed probability distributions.

Our model deals with information processing in a static manner, the focus being on the axiomatization of a general criterion and its properties. But in a sense, it begs the question of how *new* information would be processed. We offer some preliminary thoughts on how

this static model could handle arrival of new information in the form of a new possible set of distributions. In particular, we show through examples that there is no logical problems with performing dynamic programming, although a general dynamic analysis remains to be done.

Next, we study the properties of our criterion concerning notions of imprecision aversion. We define a notion of *comparative imprecision aversion* with the feature that it can be completely separated from risk attitudes. Loosely speaking, we say that a decision maker b is more imprecision averse than a decision maker a if whenever a prefers to *bet* on an event when the information is given by a (precise) probability distribution rather than some imprecise information, b prefers the bet with the precise information as well. This notion captures in rather natural terms a preference for precise information, which does not require the two decision makers that are compared to have the same risk attitudes. Our result states that two decision makers can be compared according to that notion if and only if the set of revealed probability distributions of one of them is included in the other's. We also define a notion of *imprecision premium* that is consistent with this notion of comparative imprecision aversion.

In our representation theorem, we use Gilboa and Schmeidler's axiom of uncertainty aversion which states that mixing two indifferent acts can be strictly preferred to any of these acts, for hedging reasons. We can however provide a more direct way of modeling the decision maker's attitude toward imprecision, which also provides an easy way of experimentally testing the axiom. We show in particular that uncertainty aversion is implied by an axiom of aversion toward imprecision which compares the same act under two different probability-possibility sets. Aversion toward imprecision states, loosely speaking, that the decision maker always prefers to act in a setting in which he possesses more information, i.e., the decision maker is averse toward a "garbling" of the available information. At

5

this stage, we simply remark that the notion we adopt of what it means for a probability-possibility set to be more imprecise than another one is rather weak and partial in the sense that it does not enable one to compare many sets (this will be discussed in Section 3.) One of the advantages of our setting is that it allows a clean separation between imprecision neutrality and the absence of imprecision. The latter is a feature of the information the decision maker possesses, while imprecision neutrality is characterized by the fact that the decision maker's set of revealed probability distributions is reduced to a singleton.

The next step in the paper is to provide a more specific functional form. This is done under extra axioms that capture some invariance properties, which will be discussed in Section 4. The set of revealed probability distributions 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* of the set (see Schneider (1993)). For cores of belief functions for instance, it coincides with the Shapley value of the belief function. Denoting $s(P)$ the Steiner point of P, we obtain that $\varphi(P) = (1 - \varepsilon)s(P) + \varepsilon P$ and hence (P, f) is preferred to (Q, g) if and only if

$$\varepsilon \min_{p \in P} \int u \circ f dp + (1 - \varepsilon) \int u \circ f ds(P) \geq \varepsilon \min_{p \in Q} \int u \circ g dp + (1 - \varepsilon) \int u \circ g ds(Q).$$

.

This functional form, already suggested in Ellsberg (1961), consists of taking the convex combination of the minimum expected utility with respect to all the probability-possibility set, with the expected utility with respect to a particular probability distribution in this set. The parameter ε is obtained as part of the representation result and can be interpreted as a degree of imprecision aversion. When $\varepsilon = 0$, we obtain expected utility. When $\varepsilon = 1$ the functional form expresses the extreme case where the decision maker takes the worst case scenario in the entire probability-possibility set.

A portfolio choice example illustrates the mechanics of the model. In particular, it shows how the distinction between imprecision and attitude toward imprecision can give rise to different comparative statics exercises.

Relationship with the literature

Our model incorporates information as an object on which the decision maker has well defined preferences. To the best of our knowledge, Jaffray (1989) is the first to axiomatize a decision criterion that takes into account "objective information" in a setting that is more general than risk. In his model, preferences are defined over belief functions. The criterion he axiomatizes is a weighted sum of the minimum and of the maximum expected utility. This criterion prevents a decision maker from behaving as an expected utility maximizer, contrary to ours, which obtains as a limit case the expected utility criterion. Interest in this approach has been renewed recently, in a series of papers that have in common that objects of choice are sets of lotteries. Olszewski (2007) also characterizes a version of the α-maxmin expected utility in which the decision maker puts weights both on the best-case and the worst-case scenarios. Stinchcombe (2003) characterizes a general class of expected utility for sets of lotteries. Ahn (2005) also studies preferences over sets of lotteries, and characterizes a conditional subjective expected utility in which the decision maker has a prior probability over lotteries and updates it according to each objective set. A limitation of his analysis is that the sets of probabilities considered are "regular", i.e., have the same dimension as the simplex on the space of outcome.

Notice that our model *does not* reduce to one of choice over sets of lotteries. In particular we do not impose that the betting behavior in an urn filled with 100 balls that could be black or white in unknown proportion is similar to the betting behavior in an urn filled with 100 balls that could be black, white, blue, green, or red in unknown proportion. Now, if we

were to impose that the decision makers only takes into account the induced distributions on outcomes when evaluating a pair (information,act), our framework would be comparable with, say, Olszewski (2007). Our criterion is different from his as we retain an axiom of uncertainty aversion (first in the form of uncertainty aversion as in Gilboa and Schmeidler (1989), then as a specific attitude toward "imprecise information") while he uses an axiom of Set Betweenness. Set Betweenness is in general not satisfied by our criterion. Note further that we obtain expected utility as a benchmark for imprecision neutral decision makers whereas Olszewski (2007) only studies comparative attitude toward objective ambiguity.[1]

Closely related to our analysis is Wang (2003). In his approach the available information is also explicitly incorporated in the decision model. Information takes the form of a set of probability distributions together with an anchor, i.e., a probability distribution that has particular salience. As in our analysis, he assumes that decision makers have preferences over couples (information,act). However, his axiom of ambiguity aversion is much stronger than ours and forces the decision maker to be a maximizer of the minimum expected utility taken over the entire information set. There is no scope in his model for less extreme attitude toward ambiguity. Following Wang's approach, Gajdos, Tallon, and Vergnaud (2004) proposed a weaker version of aversion toward imprecision still assuming that information was coming as a set of distributions together with an anchor. The notion of aversion toward imprecision developed in Section 3 is based on the one analyzed in Gajdos, Tallon, and Vergnaud (2004).

The notion of comparative imprecision aversion could itself be compared to the one found in Epstein (1999) and Ghirardato and Marinacci (2002). The latter define comparative ambiguity aversion using constant acts. They therefore need to control for risk

---

[1]Note that in Olszewski (2007) criterion, a weight 1/2 on the maximum and on the minimum does not correspond to a well-identified notion of ambiguity neutrality. As in Jaffray (1989), expected utility is not a particular case of his model. We get back to this issue in section 3.4.

attitudes in a separate manner and in the end, can compare (with respect to their ambi-guity attitudes) only decision makers that have the same utility functions.[2] Epstein (1999) uses in place of our bets in the definition of comparative uncertainty aversion, acts that are measurable with respect to an exogenously defined set of unambiguous events. As a consequence, in order to be compared, preferences of two decision makers have to coincide on the set of unambiguous events. If the latter is rich enough, utility functions then co-incide. Our notion of comparative imprecision aversion, based on the comparison of bets under precise and imprecise information does not require utility functions to be the same when comparing two decision makers. Said differently, risk attitudes are simply irrelevant to the imprecision aversion comparison.

The functional form axiomatized in Section 4 appears in some previous work (Gajdos, Tallon, and Vergnaud (2004) and Tapking (2004)), based on a rather different set of axioms and in a more limited setting. Kopylov (2006) also axiomatizes this functional form, for a fixed information-possibility set. In a setting similar to ours, Giraud (2006) axiomatizes a model in which the decision maker has non additive second order beliefs.

Finally, we compare our approach with Klibanoff, Marinacci, and Mukerji (2005). They provide a fully subjective model of ambiguity aversion, in which attitude toward ambiguity is captured by a smooth function over the expected utilities associated with a set of priors. The latter, as in Gilboa and Schmeidler (1989) is subjective. Hence, although their model allows for a flexible and explicit modeling of ambiguity attitudes, there is no link between the set of priors and the available information. Interestingly, part of Klibanoff, Marinacci, and Mukerji (2005)'s motivation is similar to ours, that is disentangling ambiguity attitude from the information the decision maker has. Formally, however, this separation holds in their model only if one makes the extra assumption that revealed beliefs coincide with the

---

[2]They actually mention that if one wants to compare two decision makers with different utility functions, one has first to completely elicit them.

objective information available. In particular, comparative statics are more transparent in our model, as information can be exogenously changed.

2 Extended multiple-priors model

We start with a benchmark model that extends the multiple-priors model by Gilboa and Schmeidler (1989) into the variable information setting.

2.1 Representation theorem

Let $\Omega = \mathbb{N}$ be the countable set of all the *potential* states of the world.[3] Let $\mathcal{S}$ be the family of nonempty and finite subsets of Ω. For each $S \in \mathcal{S}$, denote the set of probability measures over S by $\Delta(S)$. Let $\mathcal{P}(S)$ be the family of compact and convex subsets of $\Delta(S)$, where compactness is defined with regard to the Euclidian space $\mathbb{R}^S$. Let $\mathcal{P}$ be the family of *probability-possibility sets*, that is defined by $\mathcal{P} = \bigcup_{S \in \mathcal{S}} \mathcal{P}(S)$. For each $P \in \mathcal{P}$, its support is denoted by $supp(P)$.[4]

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

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

The set of pure outcomes is denoted by X. Let $\Delta^*(X)$ be the set of simple lotteries (probability measures with finite supports) over X. Let $\mathcal{F} = \{f : \Omega \to \Delta^*(X)\}$ be the set

---

[3]We assume that the state space is countably infinite. This assumption is not needed for Theorems 1 and 7 which remain true in a finite setting. We use it in Section 3 to prove Theorems 2, and 4 to 6. As we explain there, we could actually do the analysis in a finite setting, at the cost of cumbersome notation, and chose rather to use the modeling device of an infinite state space.

[4]For the sake of notational simplicity, we will consider that all probability distributions are defined over Ω and that $p(\omega) = 0 \ \forall \omega \in \Omega \setminus supp(p)$.

of *lottery acts*. Without loss, any lottery is viewed as a constant act which delivers that lottery regardless of states.

The domain of objects of choice is $\mathcal{P} \times \mathcal{F}$. The decision maker has a preference relation over $\mathcal{P} \times \mathcal{F}$, which is denoted by $\succsim$. The decision maker compares pairs of probability-possibility sets and acts. When

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

the decision maker 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. For $E \in \mathcal{S}$, denote $f_E g$ the act that yields $f(\omega)$ if $\omega \in E$ and $g(\omega)$ if not.

We now introduce the axioms. The first two axioms are quite standard.

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

Axiom 2 (Act Continuity) For every $P \in \mathcal{P}$ and $f, g, h \in \mathcal{F}$, if $(P, f) \succ (P, g) \succ (P, h)$, then there exist α and β in $(0, 1)$ such that

$$(P, \alpha f + (1 - \alpha)h) \succ (P, g) \succ (P, \beta f + (1 - \beta)h).$$

The third axiom states that the preference over lotteries is independent of information 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 in which the decision maker is indifferent between all lotteries.

Axiom 3 (Outcome Preference) (i) For every $P, Q \in \mathcal{P}$ and $l \in \Delta^*(X)$, $(P, l) \sim (Q, l)$, (ii) there exist $P \in \mathcal{P}$ and $l, m \in \Delta^*(X)$ such that $(P, l) \succ (P, m)$.

The following three axioms are parallel to those in Gilboa and Schmeidler (1989).

Axiom 4 (c-Independence) For every $f, g \in \mathcal{F}, l \in \Delta^*(X)$, $\lambda \in (0,1)$ and $P \in \mathcal{P}$,

$$(P, f) \succsim (P, g) \implies (P, \lambda f + (1-\lambda)l) \succsim (P, \lambda g + (1-\lambda)l).$$

Axiom 5 (Uncertainty aversion) For every $f, g \in \mathcal{F}$, $\lambda \in (0,1)$ and $P \in \mathcal{P}$,

$$(P, f) \sim (P, g) \implies (P, \lambda f + (1-\lambda)g) \succsim (P, f).$$

Axiom 6 (Monotonicity) For every $f, g \in \mathcal{F}$ and $P \in \mathcal{P}$,

$$(P, f(\omega)) \succsim (P, g(\omega)) \;\; \forall \omega \in supp(P) \implies (P, f) \succsim (P, g).$$

Theorem 1 The preference relation $\succsim$ satisfies Axioms 1 to 6 if and only if there exist a function $U : \mathcal{P} \times \mathcal{F} \to R$ which represents $\succsim$, a mixture-linear function $u : \Delta(X) \to R$ and a mapping $\varphi : \mathcal{P} \to \mathcal{P}$ such that

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

Moreover, u is unique up to positive linear transformations and φ is unique and has the property $supp(\varphi(P)) \subset supp(P)$.

We purposely kept as close as possible to the original axioms of Gilboa and Schmeidler (1989). In particular, we kept their two key axioms (c-Independence) and (Uncertainty Aversion). We will argue later that the latter can be replaced by a more explicit representation of the agent's attitude toward imprecision of the available information.

At this stage, the only property of φ is that $supp(\varphi(P)) \subset supp(P)$, meaning that the decision maker considers the information represented by P to be credible in the sense that states that are not given any weight by any of the relevant probability distributions are simply irrelevant. Note that this implies that the support of $\varphi(P)$ is finite.

2.2 The link between P and $\varphi(P)$

In our setup $\varphi(P)$ embodies attitude toward uncertainty as well as the processing of some given information. Call it the set of revealed probability distributions. Our approach is

meant to shed some light as to what is behind this set. In the representation theorem above, P is objective information, while the function φ is subjective in the same way the utility function u is. We can however specify further the property of φ that will constrain the link between the revealed probability distributions and the initial information. This link is, in Theorem 1, rather tenuous. We do that in two steps.[5]

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

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

The next axiom states that the evaluation of an act under *precise* information –that is, when the probability-possibility set is given as a singleton– depends only on its induced distribution. Notice that we do not assume the counterpart of this for general information sets. In general, two probability-possibility set/act pairs may be differently evaluated even if they induce the same sets of distributions over outcomes.

Axiom 7 (Reduction under Precise Information) For every $\{p\} \in \mathcal{P}$ and $f \in \mathcal{F}$,

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

If one adds this axiom to the ones in Theorem 1, φ has the further property that $\varphi(\{p\}) = \{p\}$ for all $\{p\} \in \mathcal{P}$. Thus, when told a precise probabilistic information the decision maker is a vNM decision maker.

The next axiom states that if one act is preferable to another under every element of the information set, the ranking is unchanged under the whole set.

Axiom 8 (Dominance) For every $f, g \in \mathcal{F}$ and $P \in \mathcal{P}$,

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

---

[5]We state informally a number of properties of φ in this section. They are given a formal treatment in the Appendix.

Another way to interpret the axiom is by saying that if $(P, g) \succ (P, f)$, then it has to be the case that there is some (precise) information in P under which g is preferred to f. Thus, the decision maker does not contemplate information outside of the set P as relevant for comparing the two acts.

If one adds (Dominance) to Axioms 1 to 6, then Theorem 1 holds with the property that $\varphi(P) \subset co\left(\cup_{p \in P} \varphi(\{p\})\right)$. As a consequence, if one combines (Reduction under Precise Information) and (Dominance), one gets the following property that establishes that $\varphi(P)$ is a selection from P.

(Selection): $\varphi(P) \subset P$ for every $P \in \mathcal{P}$.

The (Selection) property allows the following interpretation: when told P, the decision maker is assumed to know only that the true probability lies in P. With these axioms, we therefore switch to a more "objective" setting: the decision maker "truly believes" the information he is given. On the other hand, there is still some "subjectivity" in the revealed probability distributions $\varphi(P)$, but this subjectivity is constrained by the objective information P. If P is a singleton, then the constraint imposed by the (Selection) property rules out any subjectivity in the revealed probability distributions of the agent, which have to coincide with the given singleton.

We end by noting that introducing Axioms 7 to 8 entail some redundancies : under Axioms 1 to 4, (Reduction under Precise Information) and (Dominance) imply (Monotonicity).

2.3   Examples of functional forms

Theorem 1 (even augmented with the (Selection) property) is not very specific as to which form φ could take. We now illustrate the general form given in this theorem through examples.

The first example that comes to mind is when φ is the identity mapping: $\varphi(P) = P$. This could be seen as extreme pessimism (in a formal sense explained later).

The second example, which will be axiomatized in Section 4, consists of the following:

$$\varphi(P) = (1 - \varepsilon)s(P) + \varepsilon\, P$$

where $s(P)$ is the "center" of the set P.[6] The appropriate notion of center is, for general sets, the Steiner point, that will be defined in Section 4. Here, we just notice that the criterion takes the form

$$U(P, f) = (1 - \varepsilon) \sum_{\omega \in \Omega} u(f(\omega))\, s(P)(\omega) + \varepsilon \min_{p \in P} \sum_{\omega \in \Omega} u(f(\omega))\, p(\omega)$$

that is, the convex combination between an expected utility (with respect to the center of the set P) and the minimum expected utility (with respect to distributions in P). This can be viewed as what is known in the literature as the ε-contamination case. The decision maker considers the distribution $s(P)$ as the baseline distribution and then, for robustness reasons for instance, also puts weights on the minimal expected utility according to distributions "around" $s(P)$. The parameter ε is a measure of the degree of imprecision aversion (this notion will be axiomatically defined in the next section): $\varepsilon = 1$ corresponds to extreme "pessimism" while $\varepsilon = 0$ corresponds to neutrality toward imprecision.

More generally, one could consider

$$\varphi(P) = (1 - \varepsilon)e(P) + \varepsilon\, P$$

where $e(.)$ is a mapping from $\mathcal{P}$ to $\mathcal{P}$ with the property that $e(P)$ is a singleton included in P, which gives the "reference distribution" for each set P. Again, ε is a measure of imprecision aversion and $\varepsilon = 0$ means that the agent behaves as an expected utility maximizer.

---

[6]We abuse notation slightly by not making the distinction between $s(P)$ and $\{s(P)\}$.

For the third example, define for a given P, $R_P = \{E \in \mathcal{S} | E \subset supp(P) \text{ and } \#E = \#supp(P) - 1\}$, the set of subsets of $supp(P)$ whose cardinality is the one of $supp(P) - 1$. Take

$$\varphi(P) = \left\{ q \in P | \forall E \in R_P, q(E) - \min_{p \in P} p(E) \geq \left[1 - \sum_{\omega} \min_{p \in P} p(\omega)\right] \frac{\#supp(P) - 2}{\#supp(P)} - 1 \right\}.$$

To illustrate, take $P = \Delta(\{1, 2, 3\})$ and f such that $f(1) \precsim f(2) \precsim f(3)$. Then $U(P, f) = \frac{1}{2}u(f(1)) + \frac{1}{2}u(f(2))$. This type of functional would hence correspond to first truncating the act in its upper part (that is cutting its best consequences out) and then applying an expected utility computation with equal weights on the remaining states.

Another class of examples is to take $\varphi(P)$ to be the set of maximizers (or minimizers) of some function: $\varphi(P) = argmax_{p \in P} F(p)$. One can consider for instance entropy, where $F(p) = -\sum_{\omega \in supp(p)} p(\omega) \log_2(p(\omega))$.

In some applications, it is not farfetched to assume that there is a salient probability distribution in the set P (which could be, as above, its Steiner point). In that case, $F(p)$ could be taken to be the relative entropy (see Maccheroni, Marinacci, and Rustichini (2006)) with respect to that distribution. Such types of selection might be useful in some contexts, and they are allowed by our general theorem.

2.4   Arrival of information and updating

Ours is a static model. It however treats explicitly of how an agent uses information. Thus, one could wonder how could it handle arrival of new information and what kind of updating is at work? Although we have not developed a full blown dynamic model, we can offer the following thoughts on this issue.

It is important to distinguish two broad types of information. The first type is information on the realized state. In an urn example with say, red, black and yellow balls, this means information of the type "the ball drawn is not yellow". More generally, one could

conceive of a signal structure $\sigma : \Omega \to Y$, where Y is the signal space. Our model is not meant to deal with arrival of this kind of information: if one starts from a set of distributions and is then told that a particular event is true, this new situation is not necessarily best described by a set of (updated) probability distributions. Arguably one would like to affect different weights to the various updated distributions, as the realized event usually yields some information about the plausibility of the initial distributions. Chambers and Hayashi (2005) show a general impossibility result of the following form: for any selection rule[7] when told an event E, the set of selected distributions does not coincide with the Bayesian update of the originally selected distributions. Interestingly, this problem is not specific to imprecision averse decision makers: a decision maker whose selection rule always picks one single probability distribution, and hence behaves as an expected utility maximizer, faces the same problem. Thus conditioning with respect to an event is in and by itself a problem in this model, irrespective of the decision maker's attitude toward imprecision.

The second type of information bears on the distributions themselves. It could come in two different ways: either from a noisy signal on the possible distributions (sampling for instance) or from non noisy information on the possible sets of distributions (imagine merging two data base). For sampling, again, one can not remain with simply a set of distributions: the information learnt through sampling should lead the decision maker to put more or less weights on particular distributions and cannot be accommodated by our model.[8] Now, information as a new set of probability distributions can be handled and arguably, our model is precisely designed for this: in this configuration, the information

---

[7]Our φ function can be seen, under (Reduction) and (Dominance) as a selection rule according to their terminology.

[8]To get back to the urn example of the previous paragraph, sampling would correspond to the information: ten draws (with replacement) have been made and out of these draws, the ball drawn was red 3 times, black 6 times and yellow once. Getting a non noisy information on the set of distributions would for instance correspond to the information "there are as many red balls as yellow balls".

that flows in is of the same nature as "initial" information. In this situation, when merging two (incomplete) data bases, one represented by the set of distributions P_1 and the other one by P_2, the new information is simply represented by $P_1 \cap P_2$.

When new information comes as a set of distributions, we offer tentatively the following thoughts. Imagine that the decision maker is given a set of distributions P and is told that he could expect to learn that either P_1 or P_2 is the true set of distributions, with (P_1, P_2) a partition of P. Then, in some particular cases of our general functional form, we can find aggregators that allow one to perform dynamic programming in the following sense: if the decision maker prefers, conditionally to P_1, f_1 over g_1 and, conditionally to P_2, f_2 over g_2, then he prefers ex ante the act $(f_1$ if P_1, f_2 if $P_2)$ to $(g_1$ if P_1, g_2 if $P_2)$.

Example 1 Let $\varphi(P) = P$. Then, our criterion is written

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

Then the aggregator min satisfies the properties needed for dynamic programming: defining ex ante preferences by $U([P_1, f_1; P_2, f_2]) = \min[U(P_1, f_1), U(P_2, f_2)]$ satisfies all the requisite. This setting has been thoroughly studied by Chassagnon and Vergnaud (1999). They show that, more generally, dynamic consistency is satisfied.$\diamondsuit$

Example 2 Let $P \in \mathcal{P}$. Let μ be a measure on P and assume P_1 and P_2 have positive measure. Define μ^{P_1} as follows. For $A \subset P_1$, $\mu^{P_1}(A) = \frac{\mu(P_1 \cap A)}{\mu(P_1)}$ and similarly for μ^{P_2}. Let $\bar{p} = \int_P p \mu(p) dp$, $\bar{p}_1 = \int_{P_1} p \mu^{P_1}(p) dp$, and $\bar{p}_2 = \int_{P_2} p \mu^{P_2}(p) dp$. Assume that our criterion takes the form $U(P, f) = \sum_{\omega \in \Omega} u(f(\omega)) \bar{p}(\omega)$, that is the decision maker is maximizing the expected utility with respect to the distribution $\bar{p}$. Then,

$$\sum_{\omega \in \Omega} u(f(\omega)) \, \bar{p}(\omega) = \frac{\mu(P_1)}{\mu(P)} \sum_{\omega \in \Omega} u(f(\omega)) \, \bar{p}_1(\omega) + \frac{\mu(P_1)}{\mu(P)} \sum_{\omega \in \Omega} u(f(\omega)) \, \bar{p}_2(\omega)$$

Thus, here again, due to the linearity of the criterion, dynamic programming will be satisfied: if the decision maker prefers, conditionally to P_1, f_1 over g_1 and, conditionally to P_2,

f_2 over g_2, then he prefers ex ante the act $(f_1$ if P_1, f_2 if $P_2)$ to $(g_1$ if P_1, g_2 if $P_2).\diamondsuit$

The two examples above are extreme cases: in the first, one keeps the entire family P, while in the second one reduces any family to a singleton. It is therefore an open question to know if nice dynamic properties can be satisfied for intermediate cases. Next is a rather particular example, whose goal is simply to show that there is no logical impossibility to build aggregators with those properties in this setup.

Example 3 Take $P = \Delta(\{1, 2\})$, the simplex on the first two states. Assume the decision maker expects to have information in the form of a partition (P_1, P_2) of the simplex with

$$P_1 = \{(p, 1 - p,) | p \in [0, \alpha)\}$$

and

$$P_2 = \{(p, 1 - p) | p \in [\alpha, 1]\} \text{ for some } \alpha \in (0, 1).$$

Assume finally that $\varphi(P) = \varepsilon P + (1 - \varepsilon)s(P)$. In the particular case we look at here, the Steiner point is simply the center of the interval defined by P. Hence,

$$\varphi(P) = \varepsilon\{(p, 1 - p) | p \in [0, 1]\} + (1 - \varepsilon)\left(\frac{1}{2}, \frac{1}{2}\right)$$
$$\varphi(P_1) = \varepsilon\{(p, 1 - p) | p \in [0, \alpha)\} + (1 - \varepsilon)\left(\frac{\alpha}{2}, 1 - \frac{\alpha}{2}\right)$$
$$\varphi(P_2) = \varepsilon\{(p, 1 - p) | p \in [\alpha, 1])\} + (1 - \varepsilon)\left(1 - \frac{\alpha}{2}, \frac{\alpha}{2}\right)$$

For any act f, one can compute $U(P, f)$, $U(P_1, f)$ and $U(P_2, f)$ accordingly. Define an aggregator V as follows:

$$V(x, y) = \begin{cases} (x - y)\frac{(1 - \varepsilon)\alpha}{1 - 2\varepsilon\alpha + \varepsilon} + y & \text{if} \quad x \geq y \\ (y - x)\frac{(1 - \varepsilon)(1 - \alpha)}{1 - \varepsilon + 2\alpha\varepsilon} + x & \text{if} \quad x < y \end{cases}$$

One can check that it has the following properties:

- V is increasing in its two arguments,

- for any act f, $V[U(P_1, f), U(P_2, f)] = V[U(P, f))]$,

- for any acts f_1, f_2, any constant act h, and any $\alpha \in (0, 1)$,

$$V[U(P_1, \alpha f_1 + (1-\alpha)h), U(P_2, \alpha f_1 + (1-\alpha)h)] = \alpha V[U(P_1, f_1), U(P_2, f_2)] + (1-\alpha)V[U(P, h)].$$

Hence, defining an ex ante preference relation by $V[U(P_1, f_1), U(P_2, f_2)]$, enables one to do dynamic programming in this model. Note that in the formula defining V, the case $\varepsilon = 1$ corresponds to the min operator, while the case $\varepsilon = 0$ amounts to $V(x, y) = \alpha x + (1 - \alpha)y$ and therefore corresponds to the linear aggregator of Example 2.

One can also compute a value for this type of information by computing

$$\max_{f_1, f_2 \in \mathcal{A}} V[U(P_1, f_1), U(P_2, f_2)] - \max_{f \in \mathcal{A}} U(P, f)$$

where $\mathcal{A}$ is a subset of acts. This quantity is equal to

$$V[\max_{f_1 \in \mathcal{A}} U(P_1, f_1), \max_{f_2 \in \mathcal{A}} U(P_2, f_2)] - \max_{f \in \mathcal{A}} U(P, f).$$

This equality shows that dynamic consistency holds here (the decision maker is indifferent between committing to a conditional choice before he gets new information or waiting the information and then making the optimal choice). Furthermore, this quantity is always positive, showing that information is always valuable here.$\diamondsuit$

To conclude on this issue, our model can handle arrival of information when the latter comes in the same manner as the initial information, that is, sets of probability distributions. More precisely, it is possible to find aggregator that allows one to do dynamic programming, even with agents that are averse to imprecision. Thus, we contend, our model has the feature that it can accommodate both non trivial attitude toward uncertainty and a certain consistency of behavior in a dynamic setting. The general dynamic model is beyond the scope of this paper.

3 Imprecision aversion

In this Section, we define first a comparative notion of imprecision aversion and character-ize it. We then define a notion of imprecision premium and finally consider a definition of an (incomplete) order on sets of distributions that could arguably be appropriate to model imprecision of a set of distributions and show that this notion is behind Gilboa and Schmei-dler's definition of uncertainty aversion. But before dealing with these issues, we need to introduce an axiom that links the usual notion of mixture with information possibility sets that have a particular structure.

3.1 Preliminaries: mixing

Take the usual notion of mixing, that is at the heart of Gilboa and Schmeidler's approach to uncertainty aversion: $\alpha f + (1-\alpha)g$ is the act that yields the lottery $\alpha f(\omega) + (1-\alpha)g(\omega)$ in state ω. To the extent that mixing acts is seen as being equivalent to tossing a coin, we can express an idea similar to mixing by playing on the information structure rather than on the acts.

Take the following example: consider the set of distributions given by

$$Q = \{(\alpha p, (1-\alpha)p, \alpha(1-p), (1-\alpha)(1-p)) | p \in [0,1]\}$$

where α is some number in $(0,1)$. This set has the particular feature that for any $q \in Q$, $\frac{q(1)}{q(2)} = \frac{\alpha}{1-\alpha} = \frac{q(3)}{q(4)}$. To pursue, consider now an act h together with Q. Axiom 9 below assimilates (Q,h) with the following mixture: $(\Delta(\{1,2\}), \alpha f + (1-\alpha)g)$ where $f(1) = h(1)$, $g(1) = h(2)$, $f(2) = h(3)$, $g(2) = h(4)$. This new pair represents a situation in which, when state 1 occurs –which happens with any probability $p \in [0,1]$– the decision maker is faced with the lottery $\alpha h(1) + (1-\alpha)h(2)$. In state 2, he is faced with the lottery $\alpha h(3) + (1-\alpha)h(4)$. Note again that in this operation, the relative probability of being faced with $h(1)$ compared to $h(2)$ is given by $\frac{\alpha}{1-\alpha}$.

More generally, an information/act pair $(P, \alpha f + (1 - \alpha)g)$ can be interpreted by saying that a state ω is determined, according to some unknown probability $p(\omega)$ that belongs to P. Then, once the state is realized, a roulette lottery or a coin flip takes place with odds $\alpha : (1 - \alpha)$. As illustrated in Figure 1, (Q, h) where $h(\omega) = f(\frac{\omega+1}{2})$ if ω is odd, and $h(\omega) = g(\frac{\omega}{2})$ if ω is even while $Q = \{q | \exists p \in P$ s.th. $q(\omega) = \alpha p(\frac{\omega+1}{2})$ if ω is odd and $q(\omega) = (1 - \alpha)p(\frac{\omega}{2})$ if ω is even$\}$ can then be seen as the result of "collapsing" the roulette lottery in the probability distribution that determines the state. Said differently, the state now incorporates whether the coin toss ended up heads or tails. Each state is now split in two: state ω is split into (state ω, heads) and (state ω, tails) and, conditionally on being in state ω, there is a given probability distribution that is fixed across states, according to which heads or tails is determined. The next axiom states that the decision maker sees the two objects as the same. Thus, the axiom says, the operation described above is neutral for the decision maker as it does not modify the timing of the process: uncertainty first and then risk. In spirit, this axiom is very similar to the usual reduction of compound lottery axiom.[9]

Axiom 9 (Decomposition Indifference) Let $f, g, h \in \mathcal{F}$ and $P, Q \in \mathcal{P}$. If

- $h(\omega) = f(\frac{\omega+1}{2})$ if ω is odd, and $h(\omega) = g(\frac{\omega}{2})$ if ω is even and,

- $Q = \{q | \exists p \in P$ s.th. $q(\omega) = \alpha p(\frac{\omega+1}{2})$ if ω is odd and $q(\omega) = (1 - \alpha)p(\frac{\omega}{2})$ if ω is even$\}$ for some $\alpha \in [0, 1]$

---

[9]In this axiom we need the state space to be infinite, although we could develop a model with a finite state space, in which a randomizing device would be explicitly part of the probability-possibility set. In that case, we would assume a finite state space S and append to it $\{0, 1\}^{\#S}$, representing independent objective randomizing devices. We would then consider only sets of distributions on this product space that have a specific structure, i.e., the product of a set of distributions on S and of independent distributions on $\{0, 1\}$. Some invariance properties would have to be assumed and Axiom 9 rewritten. Hence, although not necessary, assuming that the state space is infinite is, in our view, more convenient.
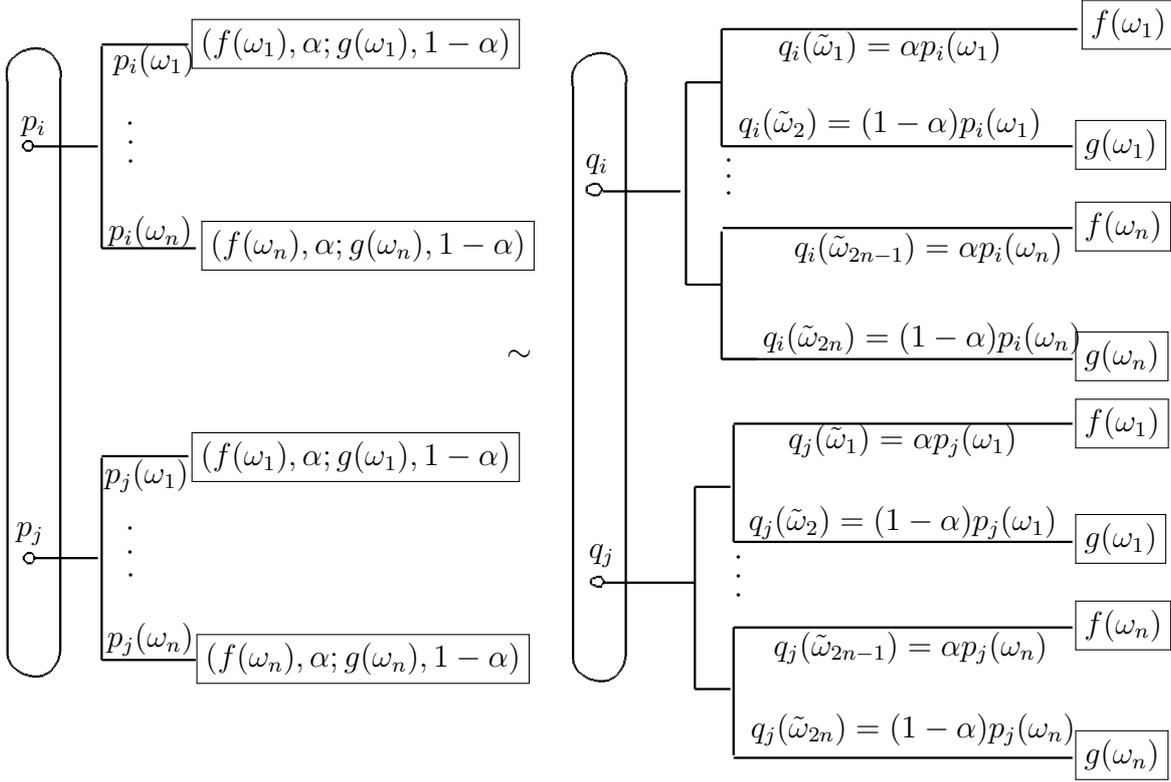
Figure 1: Decomposition Indifference

then $(P, \alpha f + (1-\alpha)g) \sim (Q, h)$.

The information embedded in the set Q is viewed as being equivalent to the information provided by the mixture operation. A natural interpretation, along the lines of the explanation given above, is to say that, for this type of information (i.e., mixing) the decision maker cares only about the induced distributions on outcomes, something reminiscent of what is imposed for precise information in Axiom 7.

(Decomposition Indifference) implies some further properties of the selection function φ.

Proposition 1 Let $\succsim$ satisfies Axioms 1 to 6, and Axiom 9. Let $P, Q \in \mathcal{P}$.
If $Q = \{q | \exists p \in P \text{ s.th. } q(\omega) = \alpha p(\frac{\omega+1}{2}) \text{ if } \omega \text{ is odd and } q(\omega) = (1-\alpha)p(\frac{\omega}{2}) \text{ if } \omega \text{ is even}\}$

for some $\alpha \in (0,1)$, then, the mapping $\varphi : \mathcal{P} \to \mathcal{P}$ in Theorem 1 has the property that

$$\varphi(Q) = \{q | \exists p \in \varphi(P) \text{ s.th. } q(\omega) = \alpha p \left(\frac{\omega + 1}{2} \right) \omega \text{ odd }, q(\omega) = (1-\alpha)p \left(\frac{\omega}{2} \right) \omega \text{ even}\}$$

$$(1)$$

Example 4 Assume that φ consists in taking the maximum entropy of P. Given that

$$\min_{p \in argmax\{-\sum_\omega p(\omega) \log_2 p(\omega)\}} \sum_\omega p(\omega)[\alpha f(\omega) + (1-\alpha)g(\omega)]$$

$$=$$

$$\min_{p \in R} \left[\sum_{\omega \text{ odd}} p(\frac{\omega+1}{2})f(\frac{\omega+1}{2}) + \sum_{\omega \text{ even}} p(\frac{\omega}{2})g(\frac{\omega}{2}) \right]$$

where $R = argmax\{-\sum_{\omega \text{ odd}} \alpha p(\frac{\omega+1}{2}) \log_2 \alpha p(\frac{\omega+1}{2}) - \sum_{\omega \text{ even}} (1-\alpha)p(\frac{\omega}{2}) \log_2 (1-\alpha)p(\frac{\omega}{2})\}$,
Axiom 9 is satisfied for this particular case.$\diamondsuit$

Example 5 Axiom 9 (and as a consequence Property (1)) will not be satisfied if, for instance $\varphi(P) = argmin_{p \in P} \sum_\omega (p(\omega) - \bar{p}(\omega))^2$ where $\bar{p}$ is some given probability distribution.$\diamondsuit$

Property (1) of the selection function will be used in all the results of this section (except Theorem 3).

3.2 Comparative imprecision aversion

One can define comparative imprecision aversion by saying that a decision maker b is more averse toward imprecision than a decision maker a if whenever he prefers an act under a singleton probability-possibility set over the same act under a general probability-possibility set, so does a. Furthermore, one would like to separate out this attitude toward imprecision from the traditional attitude toward risk. In order to do that, one has to define carefully the set of acts for which the definition applies. For two prizes $x, y \in X$ and an event E denote $x_E y$ the act giving the degenerate lottery yielding x for sure when event E realizes and the degenerate lottery yielding y for sure if E does not realize. Such an act will be called a bet in the following.

Definition 1 Let $\succsim_a$ and $\succsim_b$ be two preference relations defined on $\mathcal{P} \times \mathcal{F}$. Suppose there exist two prizes, $\bar{x}$ and $\underline{x}$ in X such that both a and b strictly prefer $\bar{x}$ to $\underline{x}$. We say that $\succsim_b$ is *more averse to bet imprecision than* $\succsim_a$ if for all $E \in \mathcal{S}$, $P \in \mathcal{P}$, and $\{p\} \in \mathcal{P}$,

$$(\{p\}, \bar{x}_E \underline{x}) \succsim_a [\succ_a](P, \bar{x}_E \underline{x}) \Rightarrow (\{p\}, \bar{x}_E \underline{x}) \succsim_b [\succ_b](P, \bar{x}_E \underline{x})$$

Theorem 2 Let $\succsim_a$ and $\succsim_b$ be two preference relations defined on $\mathcal{P} \times \mathcal{F}$ satisfying Axioms 1 to 7, and Axiom 9. Then, the following assertions are equivalent:

(i) $\succsim_b$ is more averse to bet imprecision than $\succsim_a$,

(ii) for all $P \in \mathcal{P}$, $\varphi_a(P) \subset \varphi_b(P)$.

This notion of comparative aversion to imprecision ranks preferences that do not necessarily have the same attitudes toward risk. This is of particular interest in applications if one wants to study the effects of risk aversion and imprecision aversion separately. For instance, one might want to compare portfolio choices of two agents, one being less risk averse but more imprecision averse than the other. This type of comparison cannot be done if imprecision attitudes can be compared only among preferences that have the same risk attitude, represented by the utility function. To the best of our knowledge, there is no available result in the literature that achieves this separation.

In our model, decision makers are von Neumann-Morgenstern under precise information, so that aversion toward risk is captured by the concavity of the utility index. Now, one could imagine that decision makers could distort probabilities as in rank dependent expected utility. We conjecture that Theorem 2 would remain valid in this more general setting, as the definition of comparative imprecision aversion amounts only to compare the probability of the good event, a comparison that should not be affected by increasing distortion function.

Finally, we can state a theorem similar to Theorem 2 but without assuming Axiom 9.

Theorem 3 Let $\succsim_a$ and $\succsim_b$ be two preference relations defined on $\mathcal{P} \times \mathcal{F}$ satisfying Axioms 1 to 7. Then, the following assertions are equivalent:

(i) $\succsim_b$ is more averse to bet imprecision than $\succsim_a$,

(ii) for all $P \in \mathcal{P}$,

$$\left\{ q \in P | \forall E \subset supp(P), q(E) \geq \min_{p \in \varphi_a(P)} p(E) \right\} \subset \left\{ q \in P | \forall E \subset supp(P), q(E) \geq \min_{p \in \varphi_b(P)} p(E) \right\}.$$

Thus, without Axiom 9, we do not get inclusion of the set of revealed distribution. Rather, the core of the "smallest" (convex) capacity that contains $\varphi_a(P)$ is included in the core of the "smallest" (convex) capacity that contains $\varphi_b(P)$.[10]

3.3   Imprecision premium

We define here a notion of imprecision premium which captures how much an agent is "willing to lose" when betting on an event in order to act in a setting that has no imprecision. Consider a preference relation $\succsim$ and let $\bar{x}$ and $\underline{x}$ be two prizes in X such that $\bar{x} \succ \underline{x}$. For any event $E \in \mathcal{S}$, let q^E be a probability distribution such that $(P, \bar{x}_E \underline{x}) \sim (\{q^E\}, \bar{x}_E \underline{x})$. Under Axioms 1 to 6, such a probability distribution exists and is independent of $\bar{x}$ and $\underline{x}$, since $(P, \bar{x}_E \underline{x}) \sim (\{q^E\}, \bar{x}_E \underline{x})$ if and only if $q^E(E) = \min_{p \in \varphi(P)} p(E)$. Define finally an aggregator $e : \mathcal{P} \to \mathcal{P}$ that associates with each set of distributions a particular probability distribution.[11] Examples are given by the Steiner point or (a selection of) the maximal entropy function.

Definition 2 For any $P \in \mathcal{P}(S)$, any event $E \in \mathcal{S}$, and any aggregator e, let

- the *absolute imprecision premium* with respect to $e(P)$, $\pi^A(E, P, e)$ be defined by

---

[10] $\nu(E) = \min_{p \in \varphi_a(P)} p(E)$ defines a capacity, that is a set function from the set of all subsets of $supp(P)$ to $[0, 1]$, which is monotone with respect to set inclusion. This capacity has the property of being a belief function (see Chateauneuf and Jaffray (1989) for instance).

[11] We will assimilate the singleton set $e(P)$ with the single probability distribution it contains.

$$e(P)(E) - q^E(E),$$

- the *relative imprecision premium* with respect to $e(P)$, $\pi^R(E, P, e)$ be defined by $\frac{\pi^A(E,P,e)}{e(P)(E) - Min_{p \in P} p(E)}$ whenever $e(P)(E) \neq Min_{p \in P} p(E)$.

The absolute premium is thus the mass of probability on the good event that the agent is willing to forego in order to act in a precise situation represented by $e(P)$ rather than with the imprecise probability-possibility set P. An analogy with the risk premium can be drawn as follows: $e(P)(E)$ plays the role of the expectation of the risky prospect while $q^E(E)$ plays the role of the certainty equivalent.

Theorem 4 Let $\succsim_a$ and $\succsim_b$ be two preference relations defined on $\mathcal{P} \times \mathcal{F}$, satisfying axioms Axioms 1 to 7, and Axiom 9. Let e be an aggregator function satisfying (1) in Proposition 1. Then, the following assertions are equivalent:

(i) $\succsim_b$ is more averse to bet imprecision than $\succsim_a$,

(ii) for all $P \in \mathcal{P}(S)$, and all event $E \in \mathcal{S}$, $\pi_b^A(E, P, e) \geq \pi_a^A(E, P, e)$.

Hence, as in the theory of risk aversion, one can capture by a single number the comparison of imprecision attitude. Furthermore, one can also define a notion of constant relative imprecision premium.

Definition 3 A decision maker is said to have constant relative imprecision premium θ if for any $P \in \mathcal{P}$ and any $E \in \mathcal{S}$ such that $e(P)(E) \neq Min_{p \in P} p(E)$, $\pi^R(E, P, e) = \theta$.

Theorem 5 Consider a decision maker satisfying Axioms 1 to 7, and Axiom 9. Let e be an aggregator function satisfying (1) in Proposition 1. The following assertions are equivalent:

(*i*) the decision maker has constant relative imprecision premium with respect to $e(P)$, equal to ε,

(*ii*) for all $P \in \mathcal{P}$, $\varphi(P) = \varepsilon P + (1 - \varepsilon)e(P)$.

This Theorem gives another foundation for the functional form of Theorem 7.

3.4 Imprecision aversion

In this Section, we give a new foundation for the uncertainty aversion axiom, showing that it is implied by an axiom of aversion toward imprecision. The latter compares an act under two different probability-possibility settings and states that the decision maker always prefers the more precise information. We therefore have to define a notion of imprecision on sets of probability distributions. The most natural definition would be that P is more precise than Q whenever $P \subset Q$. This is actually the definition proposed by Wang (2003). However, this definition turns out to be much too strong. Indeed, the idea behind the notion of aversion toward imprecision is that an imprecision averse decision maker should *always* prefer a more precise information, whatever the act under consideration. Consider an act f for which the worst outcome is obtained, say, in state 1. Then, Wang's notion of precision would force the decision maker to prefer $(\{(1,0)\}, f)$, that is, putting probability one on the worst outcome to $(\Delta(\{1,2\}), f)$, that is, being totally uncertain about the state; a feature of the axiom which is very unlikely and unappealing.

On the other hand, it is clear that a set being more precise than another has something to do with set inclusion. The following definition restricts the inclusion condition to some sets of probability distributions that are comparable in some sense, exactly as the comparison of two distributions in terms of risk focusses on distributions that have the same mean.

Definition 4 Let $P, Q \in \mathcal{P}$. Say that P is *conditionally more precise* than Q if

- $P \subset Q$ and,

- there exists a partition $(E_1, \ldots, E_n)$ of Ω such that

  (i) $\forall p \in P$, $\forall q \in Q$, $p(E_i) = q(E_i)$ for all $i = 1, \ldots, n$,

(ii) $co\{p(.|E_i); p \in P\} = co\{q(.|E_i); q \in Q\}$ for all $i = 1, \ldots, n$.

Note that this notion is very weak in the sense that it is very incomplete. For instance, an n-dimensional simplex cannot be compared through this definition with any of its subsets. Indeed, two sets P and Q, ordered by set inclusion, can be compared only if there exists a partition of the state space on which they agree *and* have precise probabilities (item (i) of the definition), and furthermore, conditionally on each cell of this partition, they give the same information (item (ii) of the definition). This means that the extra information contained in P is about some correlation between what happens in one cell E_i with what happens in another cell E_j. Said differently, the extra information is orthogonal to the "initial" probabilistic information, reflected in the fact that the cells of the partition have precise probabilities attached to them. The way this is expressed is via conditional probabilities: however, it should be underlined that these are simply a means to express properties of the probability-possibility sets that are compared. The use of conditional probabilities in this definition is not linked to any subjective assessment of the decision maker. Theorem 6 below provides the link between attitude toward this type of information and attitude toward hedging via mixing which is the basis for Gilboa and Schmeidler's Uncertainty Aversion axiom.

Example 6 Consider the family

$$P_\alpha = \left\{ \left(p, \frac{1}{2} - p, q, \frac{1}{2} - q \right) | p \in \left[0, \frac{1}{2} \right], q \in \left[0, \frac{1}{2} \right], |q - p| \leq \alpha \right\}$$

where $\alpha \in [0, \frac{1}{2}]$. One obviously has $P_\alpha \subset P_{\alpha'}$ for all $\alpha' \geq \alpha$. Now fix $\alpha < \frac{1}{2}$ and let $Q \equiv P_{1/2}$. $\{\{1, 2\}, \{3, 4\}, \{5, \ldots\}\}$ is a partition of the state space such that $\forall p \in P_\alpha$, $\forall q \in Q$, $p(E_i) = q(E_i)$. The set of probabilities conditional on $\{1, 2\}$ is the same when computed starting from P_α and from Q. The same is true for conditionals with respect to $\{3, 4\}$. Thus, P_α is conditionally more precise than Q. The nature of the extra information

that is present in P_α is maybe clearest for $\alpha = 0$. In that case, one has $q = p$ and the extra information that is present in P_0 is a strong correlation between the different cells of the partition. More generally, we can look at upper and lower probabilities for events according to P_α and Q. We know they agree on the partition $\{\{1,2\},\{3,4\},\{5,\dots\}\}$. One can also check that the upper and lower probabilities on the events $\{1,3\}$ and $\{2,4\}$ are the same for the two sets (0 and 1 respectively). However, the lower and upper probability of events $\{2,3\}$ and $\{1,4\}$ do differ for the two sets. One has, with obvious notation, $\underline{p}_\alpha(\{2,3\}) = 1/2 - \alpha$ and $\bar{p}_\alpha(\{2,3\}) = 1/2 + \alpha$ while $\underline{q}(\{2,3\}) = 0$ and $\bar{q}(\{2,3\}) = 1$, and similarly for event $\{1,4\}$. The fact that $\underline{p}_\alpha > \underline{q}$ and $\bar{p}_\alpha < \bar{q}$, is another way to see that P_α is more precise than $Q.\diamondsuit$

We can now state our axiom.

Axiom 10 (Aversion toward Imprecision) Let $P, Q \in \mathcal{P}$ be such that P is conditionally more precise than Q, then for all $f \in \mathcal{F}$, $(P, f) \succsim (Q, f)$.

Remark 1 Assume Theorem 1 and (Aversion toward Imprecision) hold. Then, for any $P, Q \in \mathcal{P}$ such that P is conditionally more precise than Q, $\varphi(P) \subseteq \varphi(Q)$.

For the following theorem, we need a weak form of information independence (a stronger form will be introduced and discussed in the next Section.)

Axiom 11 (Weak Information Independence) For every $P, Q \in \mathcal{P}$, $f \in \mathcal{F}$, and $\lambda \in (0,1)$,

$$(P, f) \sim (Q, f) \implies (\lambda P + (1 - \lambda)Q, f) \sim (P, f).$$

Theorem 6 Under (Weak order), (Decomposition Indifference) and (Weak Information Independence), (Aversion toward Imprecision) implies (Uncertainty Aversion).

Thus, through this theorem, we identify characteristics of the objective information that the decision maker dislikes when he prefers the mixture of two indifferent acts to either

act. The following example shows that equivalence does not hold in the previous theorem.

Example 7 Let $\alpha \in (0,1)$ and consider

$$Q = \{(\alpha p, (1-\alpha)p, \alpha(1-p), (1-\alpha)(1-p); p \in [0,1]\}$$

and

$$Q' = \{(\alpha p, (1-\alpha)q, \alpha(1-p), (1-\alpha)(1-q); p, q \in [0,1]\} = \alpha\Delta(\{1,3\}) + (1-\alpha)\Delta(\{2,4\})$$

and let $U(P,f) = \min_{p \in \varphi(P)} \sum_{\omega \in \Omega} u(f(\omega)) \; p(\omega)$. By construction, this satisfies (Uncertainty Aversion). Furthermore, it can be checked that Q is conditionally more precise than Q'. Hence, if $\varphi(Q)$ is not included in $\varphi(Q')$, (Aversion toward Imprecision) is violated.

By (Decomposition Indifference) and Proposition 1,

$$\varphi(Q) = \{(\alpha p, (1-\alpha)p, \alpha(1-p), (1-\alpha)(1-p); (p, 1-p) \in \varphi(\Delta(\{1,2\}))\}$$

Assuming (with slight abuse of notation) $\varphi(\Delta(\{1,3\})) = (\frac{1}{2}, 0, \frac{1}{2}, 0)$ and $\varphi(\Delta(\{2,4\})) = (0, \frac{1}{2}, 0, \frac{1}{2})$, one gets by (Weak Independence) that $\varphi(Q') = (\alpha\frac{1}{2}, (1-\alpha)\frac{1}{2}, \alpha\frac{1}{2}, (1-\alpha)\frac{1}{2})$. Assuming $\varphi(\Delta(\{1,2\})) = \Delta(\{1,2\}$ yields the violation of (Aversion toward Imprecision).$\diamondsuit$

Finally, we comment briefly on the link with Olszewski (2007) and α-maxmin functionals. Introduce in our model a strong notion of invariance that states that (P, f) is evaluated through the set of induced distributions on outcomes. Take the latter to be real numbers and consider $f = (1, 0, \dots)$. Then $(\Delta(\{1,2\}), f) \sim (\Delta(\{1,2,3\}), f)$. An imprecision neutral decision maker cannot, in our setting, satisfies this indifference. Indeed, for such a decision maker, $(\Delta(\{1,2\}), f) \sim (\{(\frac{1}{2}, \frac{1}{2})\}, f)$ and $(\Delta(\{1,2,3\}), f) \sim (\{(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})\}, f)$ but $(\{(\frac{1}{2}, \frac{1}{2})\}, f) \succ (\{(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})\}, f)$. On the other hand, with the α-maxmin functional $U(P, f) = \alpha \max_{p \in P} E_p[u \circ f] + (1-\alpha) \min_{p \in P} E_p[u \circ f]$, one cannot have that $(\Delta(\{1,2\}), f) \sim (\{(\frac{1}{2}, \frac{1}{2})\}, f)$ and $(\Delta(\{1,2,3\}), f) \sim (\{(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})\}, f)$: the first indifference would require $\alpha = \frac{1}{2}$ while the second would require $\alpha = \frac{1}{3}$.

4 Contraction representation: Axiomatic foundation

In this Section we provide an axiomatic characterization of the contraction representation. The contraction representation is characterized by three additional axioms: an independence axiom on mixtures of probability-possibility sets, an invariance axiom with regard to certain transformations of probability measures, and a stronger continuity axiom.

4.1 Information independence

We start by introducing the independence axiom. It states that the ranking of probability-possibility sets given an act is unchanged under taking mixtures with a common probability-possibility set. It is a natural extension of von-Neumann-Morgenstern's independence axiom to the setting of imprecise information.

Axiom 12 (Information Independence) For every $P, P', Q \in \mathcal{P}$, $f \in \mathcal{F}$, and $\lambda \in (0, 1)$,

$$(P, f) \sim (P', f) \implies (\lambda P + (1 - \lambda)Q, f) \sim (\lambda P' + (1 - \lambda)Q, f).$$

This axiom seems related to the c-Independence axiom. However, it might be worth mentioning that (Information Independence) and (c-Independence) are orthogonal to each other.[12]

To interpret, consider an object '$\lambda \circ P + (1 - \lambda) \circ Q$'. Given this, the decision maker knows that 'P is true with probability λ and Q is true with probability $1 - \lambda$.' Suppose the decision maker is to evaluate the two objects $\lambda \circ P + (1 - \lambda) \circ Q$ and $\lambda P' \circ + (1 - \lambda) \circ Q$

---

[12]A version of α-maximin representation in the form

$$U(P, f) = \alpha(f) \max_{p \in P} E_p[u \circ f] + (1 - \alpha(f)) \min_{p \in P} E_p[u \circ f]$$

where $\alpha(\cdot)$ is not constant in c-mixtures satisfies (Information Independence) but violates (c-Independence). Representation as in Theorem 1 where φ is not mixture-linear satisfies (c-Independence) but violates (Information Independence).

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 information set Q being true 'with probability $1 - \lambda$' should not matter.

We should note, however, that it is not an immediate application of the von-Neumann-Morgenstern argument. We are assuming more about indifference to the timing of resolution of uncertainty, than in the precise information case. Below we explain the types of timing indifference we are assuming with this independence axiom, so that the decision maker identifies the object $\lambda \circ P + (1 - \lambda) \circ Q$ with $\lambda P + (1 - \lambda)Q$. The argument consists of two steps.

First, the decision maker 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 decision maker 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 there is no timing effect in the sense that the decision maker is indifferent in the order of these two uncertainties (see the left half of Figure 2). This part is saying more than what the standard von-Neumann-Morgenstern-type independence axiom is assuming.

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 decision maker 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 von-Neumann-Morgenstern independence axiom over probability measures (see the right half of Figure 2).
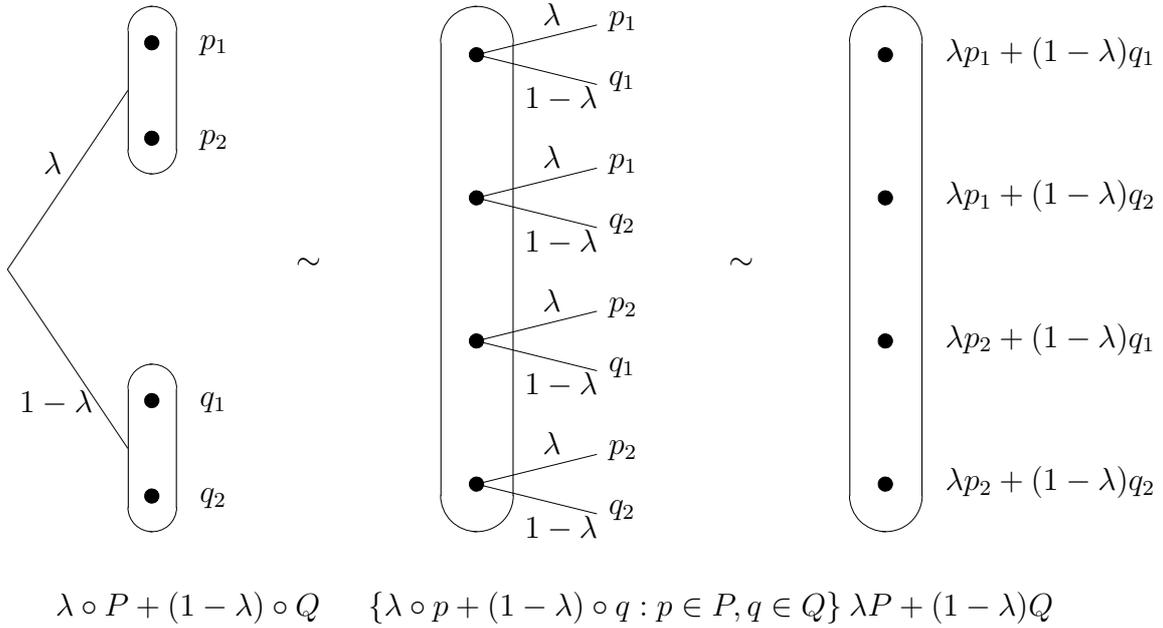
$$\lambda \Big\{ \begin{matrix} \bullet\ p_1 \\ \bullet\ p_2 \end{matrix} \qquad \sim \qquad \begin{matrix} \bullet \overset{\lambda\ p_1}{\underset{1-\lambda\ q_1}{}} \\ \bullet \overset{\lambda\ p_1}{\underset{1-\lambda\ q_2}{}} \\ \bullet \overset{\lambda\ p_2}{\underset{1-\lambda\ q_1}{}} \\ \bullet \overset{\lambda\ p_2}{\underset{1-\lambda\ q_2}{}} \end{matrix} \qquad \sim \qquad \begin{matrix} \bullet\ \lambda p_1 + (1-\lambda)q_1 \\ \bullet\ \lambda p_1 + (1-\lambda)q_2 \\ \bullet\ \lambda p_2 + (1-\lambda)q_1 \\ \bullet\ \lambda p_2 + (1-\lambda)q_2 \end{matrix}$$

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

Figure 2: Information Independence

This is how we deduce that the decision maker 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. Hence we should acknowledge that the independence axiom on mixtures of sets involves a strong timing-indifference assumption. One might think of a milder independence axiom by partially giving up the timing indifference assumptions, in particular the first one. For example, if one selects the center of gravity of a given set as her prior or base-prior, the corresponding representation violates Information Independence, while it satisfies an independence axiom with regard to mixtures of sets and singleton points (see Ahn (2005) for a related discussion).

Addition of (Information Independence) restricts our φ function to be linear in mixtures of probability-possibility sets.

Corollary 1 The preference relation $\succsim$ satisfies Axioms 1 to 6, and 12 if and only if we have the representation as in Theorem 1 with the additional property,

(Mixture-linearity): $\varphi(\lambda P + (1-\lambda)Q) = \lambda\varphi(P) + (1-\lambda)\varphi(Q) \ \forall P, Q \in \mathcal{P}, \quad \forall \lambda \in [0,1]$.

4.2　Invariance and continuity

Next we introduce the invariance axiom. It roughly says that the decision maker's attitude toward information should not change under transformations of the state space (and probability simplex) that do not change attitude toward any precise information. This is interpreted as a requirement for a sophisticated attitude toward imprecise information.

First we give an informal presentation of the invariance axiom. For each $S \in \mathcal{S}$, let $\psi : \Delta(S) \to \Delta(S)$ denote a transformation on the simplex, and let $\widetilde{\psi} : \mathcal{F} \to \mathcal{F}$ be the transformation of acts associated with ψ. We will consider a class of transformations that do not change the ranking of precise information. That is, the transformations to be considered ψ should satisfy

$$(\{p\}, f) \succsim (\{q\}, f) \Rightarrow (\{\psi(p)\}, \widetilde{\psi}(f)) \succsim (\{\psi(q)\}, \widetilde{\psi}(f)).$$

for every $p, q \in \Delta(S)$ and $f \in \mathcal{F}$.

For the transformations that satisfy the above property, we will impose an axiom in the form that for every $S \in \mathcal{S}$, every $P, Q \in \mathcal{P}(S)$ and $f \in \mathcal{F}$,

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

Below we formally identify the class of transformations that is appropriate for our argument. Here we restrict attention to a class of bistochastic matrices, that are stochastic generalization of permutations. An $|S| \times |S|$-matrix Π is *S-bistochastic* if it is nonnegative and $\sum_{\omega \in S} \Pi_{\omega\omega'} = 1$ for each $\omega' \in S$, and $\sum_{\omega' \in S} \Pi_{\omega\omega'} = 1$ for each $\omega \in S$. For an S-bistochastic matrix Π and $f \in \mathcal{F}$, define the transformed act $\Pi f \in \mathcal{F}$ by $(\Pi f)(\omega) = \sum_{\omega' \in S} \Pi_{\omega\omega'} f(\omega')$ for each $\omega \in S$, and $(\Pi f)(\omega) = f(\omega)$ for each $\omega \notin S$.

Any bistochastic matrix may be expressed as a convex combination of permutation matrices (see Birkoff (1946)). In that sense, it is a stochastic generalization of permutation. We consider a subclass of bistochastic matrices that do not change attitude toward any precise information.

Definition 5 A bistochastic transformation Π is *S-unitary* if for every $p, q \in \Delta(S)$ and $f \in \mathcal{F}$,

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

Denote the set of all S-unitary transformations by $\mathcal{T}(S)$.

We note that unitary transformations include permutations as a special case.

The following lemma shows that the class of unitary transformations is non-empty, and is a natural correspondence of the standard unitary transformation on the Euclidian space, in the probability simplex.

Lemma 1 Assume Axioms 1 to 3, 7, and 12. Then, any bistochastic transformation Π is S-unitary if and only if there exists $\lambda \in [0, 1]$ such that

$$\Pi^t \Pi = \lambda I + \frac{1 - \lambda}{|S|} E,$$

where I denotes the identity matrix and E denotes the matrix with all entries being 1.

We state the axiom.

Axiom 13 (Invariance to Unitary Transformations) For every $S \in \mathcal{S}$, any $\Pi \in \mathcal{T}(S)$, $f \in \mathcal{F}$ and $P, Q \in \mathcal{P}$,

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

To interpret, suppose that the decision maker prefers P to Q given f. Then under Axioms 1 to 3, 7, and 12, Πf induces the same ranking of probabilities as f. That is, for

any $p \in P$ and $q \in Q$,

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

and

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

Thus after the transformation, given Πf, ΠP and ΠQ play the same roles as P and Q do in the original situation given f. Therefore, it is intuitive that the ranking of information sets is unchanged, which leads to the ranking $(\Pi P, \Pi f) \succsim (\Pi Q, \Pi f)$.

The implication of the above invariance axiom is that the decision maker does not consider that a probability measure at a certain location or direction is more likely to be true. At the level of how the decision maker selects the probability distributions, it says that the selection depends only on the shape of the set, and is independent of its location and direction. We should note that the notion of unitary invariance presumes that preference follows the expected utility hypothesis under precise information, which is characterized either by the conjunction of (Reduction under Precise Information) and (c-Independence), or by the conjunction of (Reduction under Precise Information) and (Information Independence). Unitary transformation is a stochastic extension of permutation, and it is assumed that attitude toward precise information is neutral to such stochastic extensions. This is the ground assumption in which the unitary-invariance axiom makes sense.

Finally, we consider a continuity axiom with regard to probability-possibility sets.[13] Recall that for each $S \in \mathcal{S}$, $\Delta(S)$ is a compact subset of the Euclidian space $\mathbb{R}^{|S|}$, and $\mathcal{P}(S)$ is a compact metric space with regard to the Hausdorff metric.

Axiom 14 (Information Continuity): For every $S \in \mathcal{S}$, $f \in \mathcal{F}$ and $P \in \mathcal{P}(S)$, the sets

---

[13]This axiom has been criticized by Olszewski (2007). However, his criticism does not apply here as we only consider convex sets.

$\{Q \in \mathcal{P}(S) : (Q, f) \succsim (P, f)\}$ and $\{Q \in \mathcal{P}(S) : (P, f) \succsim (Q, f)\}$ are closed with regard to the Hausdorff metric.

Examples below illustrates the roles of the above axioms.

Example 8 Fix $S \in \mathcal{S}$, and let $\mathcal{P}^*(S)$ be the set consisting of *full-dimensional* compact convex subsets, and singleton points of $\Delta(S)$. Let α be the uniform distribution over $\Delta(S)$. For $P \in \mathcal{P}^*(S)$ which is non-singleton, its *center of gravity* is defined by

$$c(P) = \frac{1}{\alpha(P)} \int_P p\,\alpha(dp).$$

When the set is a singleton, its center of gravity is itself. Consider a mapping $\varphi : \mathcal{P}^*(S) \to \mathcal{P}^*(S)$ defined by

$$\varphi(P) = (1 - \varepsilon)c(P) + \varepsilon P,$$

where $\varepsilon \in [0, 1]$ is a fixed parameter.

Representation as in Theorem 1 which is restricted to $\mathcal{P}^*(S) \times \mathcal{F}$ and has φ in the above form satisfies (Invariance to Unitary Transformations), since the selection depends only on the shape of the set and is invariant to location and scale. It also satisfies (Information Continuity) on $\mathcal{P}^*(S) \times \mathcal{F}$. However, it fails to satisfy (Information Independence), and also it fails to extend to $\mathcal{P}(S) \times \mathcal{F}$ so as to maintain (Information Continuity).$\diamondsuit$

Example 9 Fix $S \in \mathcal{S}$ and let $e = (\frac{1}{|S|}, \cdots, \frac{1}{|S|})$ and $V = \{v \in \mathbb{R}^S : \langle v, e \rangle = 0, \|v\| = 1\}$ be the $|S| - 2$ dimensional unit sphere orthogonal to e. Let μ be a non-atomic probability measure over V. For $P \in \mathcal{P}(S)$, its *generalized Steiner point* with respect to μ is defined by

$$s_\mu(P) = \int_V \arg\max_{p \in P} \langle p, v \rangle \, \mu(dv).$$

Consider a mapping $\varphi : \mathcal{P}(S) \to \mathcal{P}(S)$ defined by

$$\varphi(P) = (1 - \varepsilon)s_\mu(P) + \varepsilon P,$$

where $\varepsilon \in [0, 1]$ is a fixed parameter.

Representation as in Theorem 1 with φ taking the above form satisfies (Information Independence) and (Information Continuity), but in general fails to satisfy (Invariance to Unitary Transformations) unless μ is uniform. When a permutation invariance condition is imposed on φ, we obtain that μ is permutation-invariant, i.e., $\mu(\pi \circ E) = \mu(E)$ for every Borel subset E of V and every permutation π. However, the class of permutation-invariant measures is still very large.

Whether we obtain the above class of selection mappings if we drop or weaken (Invariance to Unitary Transformations) is an open problem.$\diamondsuit$

4.3 Contraction representation result

We now provide the contraction representation in which the set revealed of probability distributions is obtained by (i) solving for the 'center' of the probability-possibility set, and (ii) shrinking the set toward the center to a degree given by preferences. The 'center' is the *Steiner point*. Imagine that a vector v is drawn from the unit sphere around the origin according to the uniform distribution. Then the Steiner point of set P, denoted by $s(P)$, is the expected maximizer of pv over $p \in P$.

More formally, fix $S \in \mathcal{S}$ and let $e = (\frac{1}{|S|}, \cdots, \frac{1}{|S|})$ and $V = \{v \in \mathbb{R}^{|S|} : \langle v, e \rangle = 0, \|v\| = 1\}$ be the $|S| - 2$ dimensional unit sphere orthogonal to e. For $P \in \mathcal{P}(S)$, its Steiner point is defined by

$$s(P) = \int_V \arg\max_{p \in P} \langle p, v \rangle \nu(dv),$$

where ν is the uniform distribution over V.[14]

Example 10 Steiner point of a segment is its midpoint.

---

[14]Multiplicity of maximizers inside the integral does not matter since uniform distribution is non-atomic.

Example 11 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 12 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. This is not surprising since in the domain of convex capacities the Shapley value is the unique single-valued selection of the core that satisfies mixture independence and permutation invariance.◇

We state the contraction representation result.

Theorem 7 The preference relation $\succsim$ satisfies Axioms 1 to 5, 7, 8, and 12 to 14 if and only if we have the representation as in Theorem 1 with the additional property that for every $S \in \mathcal{S}$, and $P \in \mathcal{P}(S)$,

$$\varphi(P) = (1 - \varepsilon)s(P) + \varepsilon\, P$$

with $\varepsilon \in [0, 1]$ that is unique.

Notice that the rate ε is constant for every probability-possibility set with finite support.

Remark 2 Under the representation of Theorem 7, a decision maker b who is more averse to bet imprecision than a decision maker a will have $\varepsilon^b > \varepsilon^a$.

5 Example

We develop in this section a simple application of our analysis to portfolio choice that is similar in spirit to Klibanoff, Marinacci, and Mukerji (2005)'s. There are three assets, a,

b, and c. The following table gives the payoff matrix

s	1	2	3	4
a	k	k	k	k
b	$\bar{b}$	$\bar{b}$	1	1
c	$\bar{c}$	1	1	$\bar{c}$

We put the following restrictions on the parameters: $\bar{c} > \bar{b} > k > 1$. The information available is given by the set

$$P_\alpha = \left\{ \left(p, \frac{1}{2} - p, q, \frac{1}{2} - q \right) | p \in \left[0, \frac{1}{2} \right], q \in \left[0, \frac{1}{2} \right], |q - p| \leq \alpha \right\}$$

where $\alpha \in [0, \frac{1}{2}]$. Hence, the probability of $\{1, 2\}$ is precise, equal to $1/2$ and similarly for $\{3, 4\}$. α is a measure of how "imprecise the set is": a higher α corresponds to a higher degree of imprecision. Taken with this information, the assets have a natural interpretation: asset a is the safe asset, b is the "risky" asset as its payoffs are measurable with respect to the partition $\{\{1, 2\}, \{3, 4\}\}$, and asset c is the "imprecise" asset.

We consider a decision maker with CARA utility function $u(w) = -e^{-\gamma w}$, where γ is the coefficient of absolute risk aversion. The transformed set is given by:

$$\varphi(P_\alpha) = \left\{ \left(p, \frac{1}{2} - p, q, \frac{1}{2} - q \right) | p \in \left[\frac{1}{4} - \theta, \frac{1}{4} + \theta \right], q \in \left[\frac{1}{4} - \theta, \frac{1}{4} + \theta \right], |q - p| \leq \alpha \right\}$$

θ is the parameter of imprecision aversion, in the sense that it gives the rate of contraction for the simplex $\Delta(\{1, 2\})$. For simplicity, we assume that the constraint on the distance between p and q is the same in the transformed set as in the information set (it is easy to generalize to a constraint of the type $|q - p| \leq \beta(\alpha)$ with $\beta(.)$ increasing in α.) To make things interesting, we assume that $\theta \geq \alpha/2$, so that the constraint $|q - p| \leq \alpha$ is effective in the computation of the optimal portfolio (although see Remark 3 below.)

The decision maker has one unit of wealth that he has to allocate among the three assets. We allow for short sales. We consider successively three cases depending on which

assets are actually available, the first case being the benchmark situation of choice between the safe and the risky asset.

Case 1: choice between safe and risky asset.

This case is the usual one and one gets that $b^\star = \frac{1}{\gamma(1-\bar{b})}log\left(\frac{k-1}{\bar{b}-k}\right)$, which is naturally independent from the parameters θ and α. Under the parameter restrictions, it is easy to see that increasing risk aversion decreases holding of the risky asset.

Case 2: choice between safe and imprecise asset.

The problem to be solved here is to find the optimal amount of the imprecise asset, i.e., the solution to: $\max_c \min_{\pi \in \varphi(P_\alpha)} - \left[(\pi(1) + \pi(4))e^{-\gamma((1-c)k+c\bar{c})} + (\pi(2) + \pi(3))e^{-\gamma((1-c)k+c)}\right]$, or rewritten in terms of p and q:

$$\max_c \min_{\varphi(P_\alpha)} - \left[(p + 1/2 - q))e^{-\gamma((1-c)k+c\bar{c})} + (1/2 - p + q)e^{-\gamma((1-c)k+c)}\right]$$

As long as $c > 0$, $-e^{-\gamma((1-c)k+c\bar{c})} > -e^{-\gamma((1-c)k+c)}$ and hence the decision maker will "use" the probability in $\varphi(P_\alpha)$ that put the highest weight on the event $\{2,3\}$ and lowest weight on $\{1,4\}$. Hence, one wants to minimize $p - q$. Let therefore $q = 1/4 + \theta$ and $p = 1/4 + \theta - \alpha$.[15] Solving for the optimal solution yields

$$c^\star = \frac{1}{\gamma(\bar{c} - 1)}log\left(\frac{(\bar{c} - k)(1/2 - \alpha)}{(k - 1)(1/2 + \alpha)}\right)$$

One can check that $c^\star$ is positive as conjectured if $(k - 1)/(\bar{c} - k) < (1/2 - \alpha)/(1/2 + \alpha)$. Here, the comparative statics with respect to γ works as in the single risky asset case. What is more interesting, although intuitive, is that the imprecise asset holding is decreasing in α: an increase in imprecision of the information provided reduces the amount of asset the

---

[15]Actually, it is easy to see that this is not the only possible choice of a minimizing probability. $q = 1/4 - \theta + \alpha$ and $p = 1/4 - \theta$ would also minimize $p - q$. The optimal solution however does not depend on which one of these probability distributions is used, as the objective function depends only on $p - q$.

decision maker wants to hold. Note also that imprecise asset holding does not depend, in this example, on the imprecision aversion parameter θ (as long as $\theta \geq \alpha/2$).

Case 3: choice among all three assets.

This is the more general case and is a bit more tedious to study. Let's write u_s the utility of the portfolio in state s. As long as $b > 0$ and $c > 0$, one has that $u_1 > u_2$ and $u_4 > u_3$ and furthermore, $u_4 - u_3 > u_1 - u_2$. Hence, the minimizing probability that belongs to $\varphi(P_\alpha)$ is given by $p = 1/4 + \theta - \alpha$ and $q = 1/4 + \alpha$.

Let $K = \frac{(\bar{c}-k)(\bar{b}-1)}{(\bar{c}-\bar{b})(k-1)}$. Under our assumption, $K > 1$. Then, the optimal solution can be written:

$$
\begin{aligned}
b^\star &= \frac{1}{\gamma(\bar{b}-1)} log\left[(K-1)\frac{1/4-\theta+\alpha}{1/4+\theta}\right] \\
c^\star &= \frac{1}{\gamma(\bar{c}-1)} log\left[\frac{\bar{c}-\bar{b}}{\bar{b}-1}\left((K-1)\frac{1/4-\theta}{1/4+\theta} + \frac{1/4+\theta-\alpha}{1/4-\theta+\alpha}\right)\right]
\end{aligned}
$$

Under some further (uninteresting) restrictions on the parameters, one can check that $b^\star > 0$ and $c^\star > 0$ as conjectured when picking the minimizing probability distribution.

One can thus perform comparative statics exercises. As α increases, that is as the information available is less precise, the decision maker will hold *more* of the risky asset and *less* of the imprecise asset. Thus, there is some form of substitution among assets as imprecision increases. This suggests that the observed under diversification of decision makers' portfolio might be a consequence of how imprecision affects different assets. More specifically, consider parameter values such that $b^\star > c^\star$ (in our toy example this is the case for a large range of parameter values.) Note that if one were to ignore the effect of uncertainty on asset holding by wrongly setting $\alpha = 0$, the predicted holding of the risky asset would be lower than $b^\star$ while the predicted holding of the imprecise asset would be higher than $c^\star$, i.e., the predicted holdings would appear to be more diversified. Thus if one fails to identify which assets are affected by imprecision, one could overestimate the predicted weight of these assets in the optimal portfolio.

Finally, it is also easy to show that the holdings of the risky as well as the imprecise assets are decreasing in the risk aversion parameter γ, as well as with the imprecision aversion parameter θ. This might help explaining phenomena like the equity premium puzzle, as imprecision aversion essentially reinforces the effect of risk aversion. Interestingly, these two very tentative hints as to how to account for the under-diversification puzzle *and* the equity premium puzzle in our model are linked to two different parameters (imprecision and imprecision aversion) and could therefore be incorporated in the *same* model.

Remark 3 The comparative static exercises performed were done under the assumption that $\theta \geq \alpha/2$. If this were not the case, then one can show that the minimizing probability used to evaluate the portfolio returns does not depend on α (when looking at the choice among all three assets.) Hence, over the full range of parameters there is a discontinuity in how imprecision affects holding of the risky and imprecise assets.

Remark 4 Note that all the action in this example does not take place because of the non-differentiability introduced by the min operator, as for instance in Epstein and Wang (1994) or Mukerji and Tallon (2001). Rather, the comparative statics were done at points where, locally, the decision maker behaves like an expected utility maximizer. More precisely, in usual maxmin expected utility models, decision makers look like expected utility maximizers away from the 45 degree line and there is no sense in which one can "change the set of priors" as there is no explicit link with the available information. In our model, there is some leverage in that respect even away from the kinks, as we have a way to link changes in the set of revealed probability distributions to changes in available information and to changes in imprecision attitudes. Thus, although non smooth, our model remains tractable in applications.

Appendix

Proof of Theorem 1

We only show sufficiency. Fix $P \in \mathcal{P}$. Then, Gilboa and Schmeidler (1989) result yields that Axioms 1 to 6 hold if and only if there exists a function $U_P : \mathcal{P} \times \mathcal{F} \to R$ such that $U_P(P, f) \geq U_P(P, g)$ if and only if $(P, f) \succsim (P, g)$ and a mixture-linear function $u_P : \Delta(X) \to \mathbb{R}$ and a unique set $\bar{P} \in \mathcal{P}$ such that

$$U_P(P, f) = \min_{p \in \bar{P}} \sum_{\omega \in \Omega} u_P(f(\omega)) \, p(\omega).$$

Moreover, u_P is unique up to positive linear transformations. For each P, define $\varphi(P) = \bar{P}$ as obtained from Gilboa and Schmeidler's theorem. The latter also implies that the decision maker is an expected utility maximizer over constant acts. Axiom 3 implies that, for any $P, Q \in \mathcal{P}$, u_P and u_Q represent the same expected utility over constant acts. Hence, they can be taken so that $u_P = u_Q = u$.

To show that the representation can be extended to the entire domain $\mathcal{P} \times \mathcal{F}$, let $(P, f) \succsim (Q, g)$. Since $S(P)$ and $S(Q)$ are finite and $f(\omega)$ and $g(\omega)$ have finite support and using Axiom (3), there exist $\bar{x}$ and $\underline{x}$ in X such that for all $\omega \in S(P) \cup S(Q)$, for all $x \in Supp(f(\omega)) \cup Supp(g(\omega))$, $(P, k_{\bar{x}}) \succsim (P, k_x) \succsim (P, k_{\underline{x}})$ where $k_{\bar{x}}$ (resp. k_x and $k_{\underline{x}}$) is the constant act giving the degenerate lottery $\delta_{\bar{x}}$ (resp. δ_x and $\delta_{\underline{x}}$) yielding $\bar{x}$ (resp. x and $\underline{x}$) for sure. Hence, by Axiom 6, we know that $(P, k_{\bar{x}}) \succsim (P, f) \succsim (P, k_{\underline{x}})$ and $(P, k_{\bar{x}}) \succsim (P, g) \succsim (P, k_{\underline{x}})$

By Axioms 1 and 2, there exists λ such that $(P, f) \sim (P, \lambda k_{\bar{x}} + (1 - \lambda)k_{\underline{x}})$. Similarly, there exists μ such that $(Q, g) \sim (Q, \mu k_{\bar{x}} + (1 - \mu)k_{\underline{x}})$. Thus,

$$(P, f) \succsim (Q, g) \iff (P, \lambda k_{\bar{x}} + (1 - \lambda)k_{\underline{x}}) \succsim (Q, \mu k_{\bar{x}} + (1 - \mu)k_{\underline{x}})$$

Now, $(P, f) \sim (P, \lambda k_{\bar{x}} + (1 - \lambda)k_{\underline{x}})$ implies that $\min_{p \in \varphi(P)} \int u \circ f \, dp = u(\lambda \delta_{\bar{x}} + (1 - \lambda)\delta_{\underline{x}})$. We also have that $\min_{p \in \varphi(Q)} \int u \circ g \, dp = u(\mu \delta_{\bar{x}} + (1 - \mu)\delta_{\underline{x}})$ and $u(\lambda \delta_{\bar{x}} + (1 - \lambda)\delta_{\underline{x}}) \geq$

$u(\mu\delta_{\overline{x}} + (1-\mu)\delta_{\underline{x}})$, which implies that

$$\min_{p\in\varphi(P)} \sum_{\omega\in\Omega} u_P(f(\omega))\, p(\omega) \geq \min_{p\in\varphi(Q)} \sum_{\omega\in\Omega} u_P(g(\omega))\, p(\omega).$$

We end by proving that $supp(\varphi(P)) \subset supp(P)$. Assume to the contrary that $supp(\varphi(P))$ is not included in $supp(P)$. Then there exists $p^* \in \varphi(P)$ and such that $p^* \notin (supp(P)) = q \neq 1$. Consider $\overline{x}$ and $\underline{x}$ in X such that $u(\delta_{\overline{x}}) > u(\delta_{\underline{x}})$ and let f be defined by $f(\omega) = \delta_{\overline{x}}$ for all $\omega \in supp(P)$, $f(\omega) = \delta_{\underline{x}}$ otherwise, and g by $g(\omega) = \delta_{\overline{x}}$ for all $\omega \in \Omega$. Then,

$$\min_{p\in\varphi(P)} \sum_{\omega\in\Omega} u(f(\omega))\, p(\omega) \leq \sum_{\omega\in\Omega} u(f(\omega))\, p^*(\omega) = qu(\overline{x})+(1-q)u(\underline{x}) < u(\overline{x}) = \min_{p\in\varphi(P)} \sum_{\omega\in\Omega} u(g(\omega))\, p(\omega)$$

Hence, $(P,g) \succ (P,f)$, a violation of Axiom 6 since $g = f_{supp(P)}g$.

Proof of the properties described in section 2.2

Property 1 Assume Theorem 1. Then, under Axiom 7, $\varphi(\{p\}) = \{p\}$.

Proof: $(\{p\}, f) \sim (\{p\}, l(p, f))$ implies by Theorem 1 that $\min_{q\in\varphi\{p\}} \sum_{\omega\in\Omega} u(f(\omega))\, p(\omega) = \sum_{\omega\in\Omega} p(\omega)u(f(\omega))$. This is true for any f. By normalization, let u be such that $u(\delta_{\overline{x}}) = 1$ and $u(\delta_{\underline{x}}) = 0$ for some $\overline{x}, \underline{x} \in X$. Let f be an act that yields $(\delta_{\overline{x}}, \delta_{\underline{x}} \ldots, \delta_{\underline{x}})$, a vector of dimension $\#supp(p)$, for state in $supp(p)$ and consider permutations on the support. One then gets that $\min_{q\in\varphi\{p\}} q(\omega) = p(\omega)$ for all $\omega \in supp(\{p\})$. Hence, $\varphi(\{p\}) = \{p\}$.

Property 2 Assume Theorem 1. Then, under Axiom 8, $\varphi(P) \subset co\left(\cup_{p\in P}\varphi(\{p\})\right)$.

Proof: Let $P \in \mathcal{P}$ and assume that $\varphi(P) \not\subset co\left(\cup_{p\in P}\varphi(\{p\})\right)$, i.e., there exists $p^* \in \varphi(P)$ such that $p^* \notin co\left(\cup_{p\in P}\varphi(\{p\})\right)$. Using a separation argument, there exists a function $\phi: \Omega \to \mathbb{R}$ such that $\int \phi dp^* < \min_{p\in co\left(\cup_{p\in P}\varphi(\{p\})\right)} \int \phi dp$. Let $\overline{x}, \underline{x} \in X$ such that $u(\delta_{\overline{x}}) > u(\delta_{\underline{x}})$. Normalize u so that $u(\delta_{\overline{x}}) = 1$ and $u(\delta_{\underline{x}}) = 0$.

Since $supp(P) \cup supp\left(co\left(\cup_{p\in P}\varphi(\{p\})\right)\right)$ is a finite set, there exist numbers $m > 0$ and ℓ, such that for all $\omega \in supp(P) \cup supp\left(co\left(\cup_{p\in P}\varphi(\{p\})\right)\right)$, $m\phi(\omega) + \ell \in [0,1]$.

Construct f as follows:

$$f(\omega) = \begin{cases} (m\phi(\omega) + \ell)\delta_{\bar{x}} + (1 - (m\phi(\omega) + \ell))\delta_{\underline{x}} & \forall \omega \in supp(P) \cup supp\left(co\left(\cup_{p \in P}\varphi(\{p\})\right)\right) \\ \delta_{\underline{x}} \text{ otherwise} \end{cases}$$

and let $\beta \equiv \sum_{\omega \in \Omega} u(f(\omega))p^*(\omega) \in [0, 1]$.

Let $g = \frac{1}{2}f + \frac{1}{2}k_{\beta\delta_{\bar{x}} + (1-\beta)\delta_{\underline{x}}}$. Since for all p,

$$\sum_{\omega \in \Omega} u(k_{\beta\delta_{\bar{x}} + (1-\beta)\delta_{\underline{x}}})p(\omega) = \beta$$

one gets that $\sum_{\omega \in \Omega} u(f(\omega))p^*(\omega) = \sum_{\omega \in \Omega} u(g(\omega))p^*(\omega)$.

Now, observe that for all p, $\sum_{\omega \in \Omega} u(g(\omega))p(\omega) = \frac{1}{2}\sum_{\omega \in \Omega} u(f(\omega))p(\omega) + \frac{1}{2}\sum_{\omega \in \Omega} u(f(\omega))p^*(\omega)$.

By construction,

$$\min_{q \in \varphi(\{p\})} \sum_{\omega \in \Omega} u(f(\omega))q(\omega) \geq \min_{q \in co\left(\cup_{p \in P}\varphi(\{p\})\right)} \sum_{\omega \in \Omega} u(f(\omega))q(\omega) > \sum_{\omega \in \Omega} u(f(\omega))p^*(\omega)$$

Thus, since $\min_{q \in \varphi(\{p\})} \sum_{\omega \in \Omega} u(g(\omega))q(\omega) = \frac{1}{2}\min_{q \in \varphi(\{p\})} \sum_{\omega \in \Omega} u(f(\omega))q(\omega) + \frac{1}{2}\sum_{\omega \in \Omega} u(f(\omega))p^*(\omega)$,

we get

$$\min_{q \in \varphi(\{p\})} \sum_{\omega \in \Omega} u(f(\omega))q(\omega) > \min_{q \in \varphi(\{p\})} \sum_{\omega \in \Omega} u(g(\omega))q(\omega)$$

and therefore $(\{p\}, f) \succ (\{p\}, g)$ for all $p \in P$.

On the other hand, $\min_{q \in \varphi(P)} \sum_{\omega \in \Omega} u(f(\omega))q(\omega) \leq \sum_{\omega \in \Omega} u(f(\omega))p^*(\omega)$. Hence,

$$\min_{q \in \varphi(P)} \sum_{\omega \in \Omega} u(g(\omega))q(\omega) = \frac{1}{2}\min_{q \in \varphi(P)} \sum_{\omega \in \Omega} u(f(\omega))q(\omega) + \frac{1}{2}\sum_{\omega \in \Omega} u(f(\omega))p^*(\omega)$$

$$\geq$$

$$\min_{q \in \varphi(P)} \sum_{\omega \in \Omega} u(f(\omega))q(\omega)$$

Therefore $(P, g) \succsim (P, f)$, a contradiction.

Property 3 Assume Theorem 1. Then, under Axioms 7 and 8, $\varphi(P) \subset P$.

Proof: Follows immediately from the previous properties.

Property 4 Assume Axioms 1 to 4 hold. Then, Axioms 7 and 8 imply Axiom 6.

Proof: Take $P \in \mathcal{P}$ and $f, g \in \mathcal{F}$ such that $(P, f(\omega)) \succsim (P, g(\omega))$ for every $\omega \in supp(P)$. By Axiom 3, we can define, for fixed P, the preference over lottery outcomes $\succsim^\star$ by $l \succsim m$ if $(P, l) \succsim^\star (P, m)$. By Axioms 2 and 4, $\succsim^\star$ satisfies the vNM conditions for the existence of a mixture linear utility function u.

$$(P, f(\omega)) \succsim (P, g(\omega)) \ \ \forall \omega \in supp(P) \ \Leftrightarrow \ u(f(\omega)) \geq u(g(\omega)) \ \ \forall \omega \in supp(P)$$

$$\Leftrightarrow \ \sum_{\omega \in supp(P)} p(\omega)u(f(\omega)) \geq \sum_{\omega \in supp(P)} p(\omega)u(g(\omega)) \ \ \forall p \in P$$

$$\Leftrightarrow \ u(l(p,f)) \geq u(l(p,g)) \ \ \forall p \in P$$

$$\Leftrightarrow \ (\{p\}, f) \succsim (\{p\}, g) \ \text{ by Axiom 7}$$

$$\Leftrightarrow \ (P, f) \succsim (P, g) \ \text{ by Axiom 8.}$$

Proof of Proposition 1

Let $P, Q \in \mathcal{P}$ and $\alpha \in [0,1]$ satisfy the assumptions of the Proposition. We first prove that $\varphi(Q) \subseteq Q^*$ where

$$Q^* = \{q | \exists p \in \varphi(P) \text{ s.th. } q(\omega) = \alpha p(\frac{\omega+1}{2}) \text{ if } \omega \text{ odd and } q(\omega) = (1-\alpha)p(\frac{\omega}{2}) \text{ if } \omega \text{ even}\}$$

Assume that there exists $p^* \in \varphi(Q)$ such that $p^* \notin Q^*$. Since Q^* is a convex set, using a separation argument, there exists a function $\phi : \Omega \to \mathbb{R}$ such that $\int \phi dp^* < \min_{p \in Q^*} \int \phi dp$. Since $supp(Q)$ is a finite set, there exist numbers a, b with $a > 0$, such that $\forall \omega \in supp(Q)$, $(a\phi(\omega) + b) \in u(\Delta(X))$. Then, for all $\omega \in supp(Q)$ there exists $y(\omega) \in \Delta(X)$ such that $u(y(\omega)) = a\phi(\omega) + b$. Define $h \in \mathcal{F}$ by $h(\omega) = y(\omega)$ for all $\omega \in supp(Q)$, $h(\omega) = \delta_x$ for all $\omega \in \Omega \backslash supp(Q)$, where $x \in X$.

Then define $f, g \in \mathcal{F}$ by $f(\omega) = h(2\omega - 1)$ and $g(\omega) = h(2\omega)$. We have that $supp(Q^*) \subseteq supp(Q)$ and therefore

$$\min_{p \in Q^*} \sum_{\omega \in \Omega} u(h(\omega)) \, p(\omega) = \min_{p \in \varphi(P)} \sum_{\omega \in \Omega} u(\alpha f(\omega) + (1 - \alpha)g(w)) \, p(\omega)$$

48

while

$$\min_{p \in Q^*} \sum_{\omega \in \Omega} u(h(\omega)) \, p(\omega) > \sum_{\omega \in \Omega} u(h(\omega)) \, p^*(\omega) \geq \min_{p \in \varphi(Q)} \sum_{\omega \in \Omega} u(h(\omega)) \, p(\omega)$$

and thus

$$(P, \alpha f + (1 - \alpha)g) \succ (Q, h)$$

which contradicts Axiom 9.

The same kind of proof permits to show that $Q^* \subseteq \varphi(Q)$.

Proof of Theorem 2

$[(i) \Rightarrow (ii)]$ Let $P \in \mathcal{P}$. Assume that $\varphi_a(P) \not\subset \varphi_b(P)$, i.e., there exists $p^* \in \varphi_a(P)$ such that $p^* \notin \varphi_b(P)$. Using a separation argument, there exists a function $\phi : \Omega \to \mathbb{R}$ such that $\int \phi dp^* < \min_{p \in \varphi_b(P)} \int \phi dp$. Let $\bar{x}$ and $\underline{x}$ in X be such that both a and b strictly prefer $\bar{x}$ to $\underline{x}$. Note that we can choose by normalization u_a and u_b such that $u_a(\bar{x}) = u_b(\bar{x}) = 1 > u_a(\underline{x}) = u_b(\underline{x}) = 0$. Since $supp(P)$ is a finite set, there exist numbers $m > 0$ and ℓ, such that for all $\omega \in supp(P)$, $m\phi(\omega) + \ell \in [0, 1]$. Denote $supp(P) = \{1, .., n\}$.[16] Let $\alpha_\omega = m\phi(\omega) + \ell$, $\omega \in supp(P)$..

Let $f^0 \in \mathcal{F}$ such that $f^0(\omega) = \alpha_\omega \delta_{\bar{x}} + (1 - \alpha_\omega)\delta_{\underline{x}}$ for all $\omega = 1, ..., n$ and $f^0(\omega) = \delta_{\underline{x}}$ for all $\omega \in \Omega \backslash supp(P)$.

f^0 is an act that can be used to "separate" $p^\star$ from $\varphi_b(P)$. It has the feature that, in each state, it yields a lottery on the same outcomes $\bar{x}$ and $\underline{x}$. The rest of the proof consists in using Axiom 9 recursively to build an equivalent pair act/probability-possibility set in which the act is now a bet of the form $\bar{x}_E \underline{x}$.

At each stage of the recursion, each state ω is split into two, say $\bar{\omega}$ and $\underline{\omega}$. At each stage, build a new act by taking the appropriate state (defined by the recursion) and decomposing

---

[16]This is not without loss of generality. However, it is straightforward conceptually although notationally involved to generalize the proof to the case where $supp(P)$ to any finite set $\{\omega_1, \ldots, \omega_n\}$.
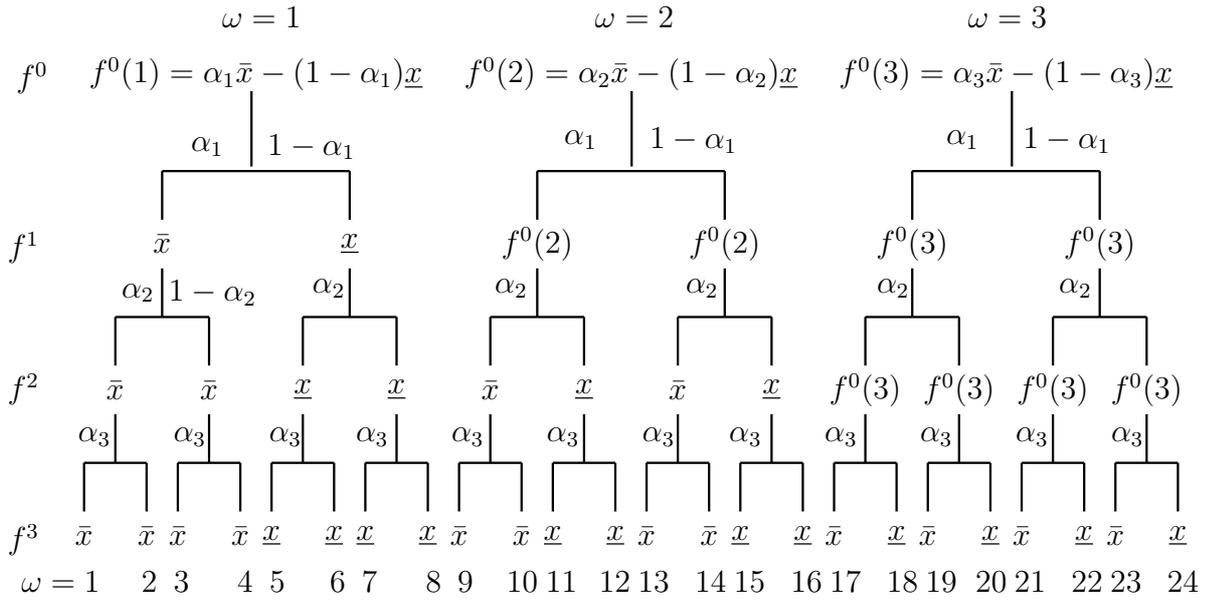
Figure 3: An example of construction of a bet

the lottery $\alpha \bar{x} + (1 - \alpha)\underline{x}$ in state ω into an act that yields $\bar{x}$ for sure in "sub-state" $\bar{\omega}$ and $\underline{x}$ in "sub-state" $\underline{\omega}$. For all the other states, simply replicate the act (i.e., the value in each sub-state is the same.)

Formally, define $f^i \in \mathcal{F}$ for $i = 1, ..., n$ as follows:

- For all $\omega \in \{1, \cdots, 2^{i-1}n\}$,

    - If $f^{i-1}(\omega) \neq f^0(i)$, then $f^i(2\omega - 1) = f^i(2\omega) = f^{i-1}(\omega)$

    - If $f^{i-1}(\omega) = f^0(i)$, then $f^i(2\omega - 1) = \delta_{\bar{x}}$ and $f^i(2\omega) = \delta_{\underline{x}}$

- For all $\omega > 2^i n$, $f^i(\omega) = \delta_{\underline{x}}$.

Figure 3 illustrates the recursion for an act f^0 whose support is $\{1, 2, 3\}$, which is ultimately spread, via duplication of the states, on a bet that involves 24 states.

A similar operation has to be done for the probability-possibility sets, so as to maintain indifference throughout.

- $Q^1 = \{q | \exists p \in P \text{ s.th. } q(2\omega - 1) = \alpha_1 p(\omega) \text{ and } q(2\omega) = (1 - \alpha_1)p(\omega) \text{ for all } \omega \in \{1, \cdots, n\}\}$

50

- for $i = 2, ..., n$, $Q^i = \{q | \exists p \in Q^{i-1} \text{ s.th. } q(2\omega - 1) = \alpha_i p(\omega) \text{ and } q(2\omega) = (1 - \alpha_i)p(\omega)$

  for all $\omega \in \{1, \cdots, 2^{i-1}n\}\}$

Finally, expand p^* in a similar way:

- p^{*1} be such that $p^{*1}(2\omega - 1) = \alpha_1 p^*(\omega)$ and $p^{*1}(2\omega) = (1 - \alpha_1)p^*(\omega)$ for all $\omega \in$

  $\{1, \cdots, n\}$

- for $i = 2, ..., n$, let p^{*i} be such that $p^{*i}(2\omega - 1) = \alpha_i p^{*i-1}(\omega)$ and $p^{*i}(2\omega) = (1 -$

  $\alpha_i)p^{*i-1}(2\omega)$ for all $\omega \in \{1, \cdots, 2^{i-1}n\}$.

By Proposition 1, one can check that:

- for $h = a, b$

$$\min_{p \in \varphi_h(P)} \sum_{\omega \in \Omega} u_h(f^0(\omega))\, p(\omega) = \min_{p \in \varphi_h(Q^1)} \sum_{\omega \in \Omega} u_h(f^1(\omega))\, p(\omega)$$

$$\text{and} \sum_{\omega \in \Omega} u_h(f^0(\omega))\, p^*(\omega) = \sum_{\omega \in \Omega} u_h(f^1(\omega))\, p^{*1}(\omega)$$

- for $h = a, b$, for $i = 2, ..., n$

$$\min_{p \in \varphi_h(Q^{i-1})} \sum_{\omega \in \Omega} u_h(f^{i-1}(\omega))\, p(\omega) = \min_{p \in \varphi_h(Q^i)} \sum_{\omega \in \Omega} u_h(f^i(\omega))\, p(\omega)$$

$$\text{and} \sum_{\omega \in \Omega} u_h(f^{i-1}(\omega))\, p^{*i-1}(\omega) = \sum_{\omega \in \Omega} u_h(f^i(\omega))\, p^{*i}(\omega)$$

- for $i = 1, ..., n$, observe that $p^{*i} \in \varphi_a(Q^i)$. Hence,

$$\min_{p \in \varphi_b(P)} \sum_{\omega \in \Omega} u_b(f^0(\omega))\, p(\omega) = \min_{p \in \varphi_b(Q^n)} \sum_{\omega \in \Omega} u_b(f^n(\omega))\, p(\omega) >$$

$$\sum_{\omega \in \Omega} u_b(f^0(\omega))\, p^*(\omega) = \sum_{\omega \in \Omega} u_b(f^n(\omega))\, p^{*n}(\omega)$$

  while

$$\sum_{\omega \in \Omega} u_a(f^0(\omega))\, p^*(\omega) = \sum_{\omega \in \Omega} u_a(f^n(\omega))\, p^{*n}(\omega) \geq$$

$$\min_{p \in \varphi_a(P)} \sum_{\omega \in \Omega} u_a(f^0(\omega))\, p(\omega) = \min_{p \in \varphi_a(Q^n)} \sum_{\omega \in \Omega} u_a(f^n(\omega))\, p(\omega)$$

Observe finally that, as a result of the way the recursion works, f^n is of the form $\bar{x}_E \underline{x}$. Now, since by (Reduction under Precise Information) $\varphi(\{p^{*n}\}) = \{p^{*n}\}$. Therefore, we have

$$(\{p^{*n}\}, \bar{x}_E \underline{x}) \succeq_a (Q^n, \bar{x}_E \underline{x})$$

while

$$(Q^n, \bar{x}_E \underline{x}) \succ_b (\{p^{*n}\}, \bar{x}_E \underline{x})$$

which contradicts the fact that $\succeq_b$ is more averse to bet imprecision than $\succeq_a$.

$[(ii) \Rightarrow (i)]$ Straightforward.

Proof of Theorem 3

$[(i) \Rightarrow (ii)]$ By (i), for any $q \in P$ and any $E \subset supp(P)$, $q(E) \geq \min\limits_{p \in \varphi_a(P)} p(E)$ imply $q(E) \geq \min\limits_{p \in \varphi_b(P)} p(E)$. Hence, $\min\limits_{p \in \varphi_a(P)} p(E) \geq \min\limits_{p \in \varphi_b(P)} p(E)$, proving (ii).

$[(ii) \Rightarrow (i)]$ Straightforward.

Proof of Theorem 4

Given the equivalence proved in Theorem 2, we show that $\varphi_a(P) \subset \varphi_b(P)$ implies (ii) and then show that (ii) implies (i).

$[\varphi_a(P) \subset \varphi_b(P) \Rightarrow (ii)]$ Since $\pi_a^A(E, P, e) = e(P)(E) - Min_{p \in \varphi_a(P)} p(E)$ and $\pi_b^A(E, P, e) = e(P)(E) - Min_{p \in \varphi_b(P)} p(E)$, $\varphi_a(P) \subset \varphi_b(P)$ implies that $\pi_b^A(E, P, e) \geq \pi_a^A(E, P, e)$.

$[(ii) \Rightarrow (i)]$ Consider prizes $\bar{x}$ and $\underline{x}$ in X such that both a and b strictly prefer $\bar{x}$ to $\underline{x}$, and let $P \in \mathcal{P}$, and $E \subset \Omega$. For any p, for any agent $i = a, b$, $(\{p\}, \bar{x}_E \underline{x}) \succsim_i [\succ_i](P, \bar{x}_E \underline{x})$ if, and only if, $\pi_i^A(E, P, e) \geq [\succ] e(P)(E) - p(E)$. Therefore since $\pi_b^A(E, P, e) \geq \pi_a^A(E, P, e)$, we have

$$(\{p\}, \bar{x}_E \underline{x}) \succsim_a [\succ_a](\bar{x}_E \underline{x}, P) \Rightarrow (\{p\}, \bar{x}_E \underline{x}) \succsim_b [\succ_b](P, \bar{x}_E \underline{x})$$

which completes the proof that $\succsim_b$ is more averse to bet imprecision than $\succsim_a$.

Proof of Theorem 5

$[(i) \Rightarrow (ii)]$ Let $P \in \mathcal{P}$, and p be a boundary point p of $co(P)$. Define:

$$\overline{\varepsilon} = Sup\left\{\varepsilon' | \varepsilon' \in [0,1] \text{ s.th. } (\varepsilon'p + (1-\varepsilon')s(P)) \in \varphi(P)\right\}.$$

Then $\overline{p} = \overline{\varepsilon}p + (1-\overline{\varepsilon})s(P)$ is a boundary point of $\varphi(P)$ since $\varphi(P)$ is closed. Since it is convex as well, there exists a function $\phi : S \to \mathbb{R}$ such that $\int \phi d\overline{p} = \min_{p \in \varphi(P)} \int \phi dp$.

Using the notation and definitions introduced in the proof of Theorem 2 in order to define $f^n = \overline{x}_E \underline{x}$, $p^n, \overline{p^n}$ and Q^n, we have that $(f^n, \{\overline{p^n}\}) \sim (f^n, Q^n)$. Note that $\overline{p^n} = \overline{\varepsilon}p^n + (1-\overline{\varepsilon})s(Q^n)$. Thus

$$\pi^R(E, Q^n) = \frac{s(Q^n)(E) - \overline{p^n}(E)}{s(Q^n)(E) - Min_{q \in Q^n}q(E)} \leq \frac{s(Q^n)(E) - \overline{p^n}(E)}{s(Q^n)(E) - p^n(E)} = \overline{\varepsilon}.$$

If $\varepsilon > \overline{\varepsilon}$ we get a contradiction with the fact that $\pi^R(E, Q^n) = \varepsilon$. Therefore, for any boundary point p of $co(P)$, $\overline{\varepsilon}(p) = Sup\{\varepsilon' | \varepsilon' \in [0,1] \text{ s.th. } (\varepsilon'p + (1-\varepsilon)s(P)) \in \varphi(P)\}$ is such that $\overline{\varepsilon}(p) \geq \varepsilon$. Let p^* be a boundary point of $co(P)$ such that $\overline{\varepsilon}(p^*) \geq \overline{\varepsilon}(p)$ for all boundary point p of $co(P)$. Then, there exists a function $\phi : S \to \mathbb{R}$ such that $\int \phi dp^* = \min_{p \in P} \int \phi dp$. Define $\overline{p^*} = \overline{\varepsilon}(p^*)p^* + (1-\overline{\varepsilon}(p^*))s(P)$ and consider now $p' \in \varphi(P)$. There exists a boundary point p of $co(P)$ and $\varepsilon' < \overline{\varepsilon}(p)$ such that $p' = \varepsilon'p + (1-\varepsilon')s(P)$.

Let us use again the notation and definition introduced in the proof of Theorem 2. Since $\int u \circ (\overline{x}_E \underline{x})\, dp^{*n} \leq \int u \circ (\overline{x}_E \underline{x})\, dp^n$ and $\int u \circ (\overline{x}_E \underline{x})\, dp^{*n} \leq \int u \circ (\overline{x}_E \underline{x})\, ds(Q^n)$, we have that $\int u \circ (\overline{x}_E \underline{x})\, d\overline{p^{*n}} \leq \int u \circ (\overline{x}_E \underline{x})\, dp'^n$. Thus $\int u \circ (\overline{x}_E \underline{x})\, d\overline{p^{*n}} = \min_{r \in \varphi(Q^n)} \int u \circ (\overline{x}_E \underline{x})\, dr$ while $\int u \circ (\overline{x}_E \underline{x})\, d\overline{p^{*n}} = \min_{r \in \varphi(Q^n)} \int u \circ (\overline{x}_E \underline{x})\, dr$. Therefore

$$\pi^R(E, Q^n) = \frac{s(Q^n)(E) - \overline{q^*}(E)}{s(Q^n)(E) - Min_{q \in Q^n}q(E)} = \overline{\varepsilon}(p^*),$$

and thus $\overline{\varepsilon}(p^*) = \varepsilon$. Hence, for all boundary point p of $co(P)$, $\overline{\varepsilon}(p) = \varepsilon$ which proves that $\varphi(P) = \varepsilon P + (1-\varepsilon)s(P)$.

$[(ii) \Rightarrow (i)]$ Consider $P \in \mathcal{P}$, and $E \subset \Omega$ such that $s(P)(E) \neq Min_{p \in P} p(E)$. We have

$$\min_{p \in \varphi(P)} p(E) = \varepsilon \min_{p \in \varphi(P)} p(E) + (1 - \varepsilon)s(P)(E),$$

and therefore

$$\pi^R(E, P) = \frac{s(P)(E) - Min_{p \in \varphi(P)} p(E)}{s(P)(E) - Min_{p \in \varphi(P)} p(E)} = \varepsilon.$$

Proof of Theorem 6

Let $f, g \in \mathcal{F}$ and $P \in \mathcal{P}$ and assume $(P, f) \sim (P, g)$. Define h by $h(\omega) = f(\frac{\omega+1}{2})$ if ω is odd, and $h(\omega) = g(\frac{\omega}{2})$ if ω is even. For any $\alpha \in [0, 1]$, define $Q(\alpha) = \{q | \exists p \in P$ s. th. $q(\omega) = \alpha p(\frac{\omega+1}{2})$ if ω is odd and $q(\omega) = (1 - \alpha)p(\frac{\omega}{2})$ if ω is even$\}$.

By (Decomposition Indifference) $(Q(1), h) \sim (P, f)$ and $(Q(0), h) \sim (P, g)$ and hence, since by assumption $(P, g) \sim (P, f)$ and $\succsim$ is transitive, $(Q(0), h) \sim (Q(1), h)$. By (Weak Information Independence) for any $\lambda \in [0, 1]$, $(\lambda Q(1) + (1 - \lambda)Q(0), h) \sim (Q(0), h)$. Hence, $(\lambda Q(1) + (1 - \lambda)Q(0), h) \sim (P, f)$.

Now, $Q(\lambda)$ is conditionally more precise than $\lambda Q(1) + (1 - \lambda)Q(0)$. Indeed, (i) $Q(\lambda) \subset \lambda Q(1) + (1 - \lambda)Q(0)$ and (ii) take as a partition of the state space $\{E_1, E_2, \dots\}$ where $E_n = \{2n - 1, 2n\}$ for $n = 1, \dots$, then the condition holds. Hence, $(Q(\lambda), h) \succ (P, f)$ and therefore, since (Decomposition Indifference) implies that $(Q(\lambda), h) \sim (P, \lambda f + (1 - \lambda)g)$, we have: $(P, \lambda f + (1 - \lambda)g) \succ (P, f)$.

Proof of Theorem 7

Mixture linearity

Lemma 2 Under (Information Independence), for every $P, Q \in \mathcal{P}$, $f \in \mathcal{F}$ and $\lambda \in [0, 1]$,

$$U(\lambda P + (1 - \lambda)Q, f) = \lambda U(P, f) + (1 - \lambda)U(Q, f).$$

Proof. Let $\{p\}, \{q\} \in \mathcal{P}$ be such that $(P, f) \sim (\{p\}, f)$ and $(Q, f) \sim (\{q\}, f)$, respectively. Then, repeated application of (Information Independence) delivers

$$(\lambda P + (1 - \lambda)Q, f) \sim (\lambda\{p\} + (1 - \lambda)Q, f) \sim (\lambda\{p\} + (1 - \lambda)\{q\}, f)$$

Since $U(\lambda\{p\} + (1 - \lambda)\{q\}, f) = \lambda U(\{p\}, f) + (1 - \lambda)U(\{q\}, f)$ is true for precise information, we obtain the claim. ∎

Lemma 3 For every $P, Q \in \mathcal{P}$ and $\lambda \in [0, 1]$, $\varphi(\lambda P + (1 - \lambda)Q) = \lambda\varphi(P) + (1 - \lambda)\varphi(Q)$.

Proof. By construction,

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

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

By mixture-linearity of U over $\mathcal{P}$, the above is equal to

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

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

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

This also serves as a proof of Corollary 1.

Continuity

For a while we fix $S \in \mathcal{S}$. We first prove continuity of φ in $P(S)$.

Lemma 4 The mapping $\varphi : \mathcal{P}(S) \to \mathcal{P}(S)$ is continuous with respect to the Hausdorff metric.

Proof. Let $\{P^n\}$ be a sequence in $\mathcal{P}(S)$ converging to $P \in \mathcal{P}(S)$. Because $\mathcal{P}(S)$ is compact, without loss of generality we 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

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

which is a contradiction to (Information Continuity). ∎

Unitary invariance

We now prove Lemma 1.

Proof. ($\Rightarrow$): Without loss of generality, we deal with payoff vectors.

In the payoff space, the unitary condition is equivalent to saying that

$$
\langle p, x \rangle \geq \langle q, x \rangle \quad \Longrightarrow \quad \langle \Pi p, \Pi x \rangle \geq \langle \Pi q, \Pi x \rangle.
$$

This implies that

$$
(p - q)^t x = 0 \quad \Longrightarrow \quad (p - q)^t \Pi^t \Pi x = 0.
$$

Take any $\omega \in S$ and let δ_ω be the vector which assigns 1 on the ω-th coordinate and 0 on the others. Then, $(p - q)^t \delta_\omega = 0$ implies $p_\omega = q_\omega$.

Now take $p, q \in \Delta(S)$ such that $(p - q)_{\omega'} = \alpha > 0$ for $\omega' \neq \omega$ and $(p - q)_{\omega''} = -\alpha$ for $\omega'' \neq \omega, \omega'$. Then,

$$
(p - q)^t \Pi^t \Pi \delta_\omega = \alpha(\Pi_{\omega'}^t \Pi_\omega - \Pi_{\omega''}^t \Pi_\omega) = 0,
$$

which implies $\Pi_{\omega'}^t \Pi_\omega = \Pi_{\omega''}^t \Pi_\omega$. Since ω is arbitrary, this implies that all the off-diagonal entries of $\Pi^t \Pi$ are the same. Therefore, all the diagonal entries of $\Pi^t \Pi$ are the same too.

Remaining to show is that the diagonal entries of $\Pi^t \Pi$ cannot be smaller than the off-diagonal entries. Denote the diagonal entry by a, off-diagonal entry by b. Let p be the vector

in $\Delta(S)$ given by $p_\omega = 1 - (|S| - 1)c$ and $p_{\omega'} = c$ for all $\omega' \neq \omega$, where $1 \leq c \leq \frac{1}{|S|-1}$. Then

we have $(\delta_\omega - p)^t \delta_\omega = (|S| - 1)c \geq 0$. The unitary condition implies that $(\delta_\omega - p)^t \Pi^t \Pi \delta_\omega \geq 0$.

Since $(\delta_\omega - p)^t \Pi^t \Pi \delta_\omega = (|S| - 1)c(a - b)$, we obtain the desired result.

($\Leftarrow$): Since

$$
\begin{aligned}
\langle \Pi p, \Pi x \rangle &= p^t \Pi^t \Pi x \\
&= p^t \left(\lambda I + \frac{1 - \lambda}{|S|} E \right) x \\
&= \lambda \langle p, x \rangle + \frac{1 - \lambda}{|S|} \langle e, x \rangle,
\end{aligned}
$$

we obtain the desired property

$$
\langle p, x \rangle \geq \langle q, x \rangle \implies \langle \Pi p, \Pi x \rangle \geq \langle \Pi q, \Pi x \rangle.
$$

$\blacksquare$

For later use, we show the lemma below

Lemma 5 For any bistochastic matrix Π, the following two statements are equivalent:

(i) The matrix $\Pi^t \Pi$ takes the form

$$
\Pi^t \Pi = \lambda I + \frac{1 - \lambda}{|S|} E,
$$

for a given $\lambda \in [0, 1]$, where I is the identity matrix and E is a matrix in which all the entries are 1.

(ii) The matrix Π satisfies $\|\Pi p - \Pi q\| = \sqrt{\lambda} \|p - q\|$ for any $p, q \in \Delta(S)$ for a given $\lambda \in [0, 1]$, where $\| \cdot \|$ denotes the Euclidian norm;

Proof. For the proof, let $e = (\frac{1}{|S|}, \cdots, \frac{1}{|S|})$. Let δ_ω be the vector which assigns 1 on the ω-th coordinate and 0 on the others. Also let Π_ω be the ω-th column vector of Π.

(i) $\Rightarrow$ (ii): It follows from

$$\langle \Pi p - \Pi q, \Pi p - \Pi q \rangle \ = \ (p-q)^t \Pi^t \Pi (p-q)$$
$$= \ (p-q)^t \left(\lambda I + \frac{1-\lambda}{|S|} E \right)(p-q)$$
$$= \ \lambda (p-q)^t I (p-q) + \frac{1-\lambda}{|S|}(p-q)^t E (p-q)$$
$$= \ \lambda \langle p-q, p-q \rangle,$$

where the last line follows from $E(p-q) = Ep - Eq = \mathbf{1} - \mathbf{1} = 0$.

(ii) $\Rightarrow$ (i): Condition (ii) is written as

$$\langle \Pi p - \Pi q, \Pi p - \Pi q \rangle = \lambda \langle p-q, p-q \rangle$$

for all $p, q \in \Delta(S)$. In particular, by taking $q = e$, we have

$$\langle \Pi p - \Pi e, \Pi p - \Pi e \rangle = \lambda \langle p-e, p-e \rangle$$

Since $\langle \Pi p, e \rangle = \langle p, e \rangle = \langle e, e \rangle = \frac{1}{|S|}$, the above condition reduces to

$$p^t \Pi^t \Pi p = \lambda p^t p + \frac{1-\lambda}{|S|}$$

By taking $p = \delta_\omega$, we have

$$\Pi_\omega^t \Pi_\omega = \lambda + \frac{1-\lambda}{|S|},$$

which is the (ω, ω) diagonal entry of $\Pi^t \Pi$. Since ω is arbitrary, all the diagonal entries are the same.

To show the claim for the off-diagonal entries, let $p = \delta_\omega$ and $q = \delta_{\omega'}$, where $\omega' \neq \omega$. Then the condition reduces to

$$\Pi_\omega^t \Pi_\omega - 2\Pi_\omega^t \Pi_{\omega'} + \Pi_{\omega'}^t \Pi_{\omega'} = 2\lambda.$$

Since $\Pi_\omega^t \Pi_\omega = \Pi_{\omega'}^t \Pi_{\omega'} = \lambda + \frac{1-\lambda}{|S|}$, we obtain

$$\Pi_\omega^t \Pi_{\omega'} = \frac{1-\lambda}{|S|}$$

which is the (ω, ω') off-diagonal entry of $\Pi^t\Pi$. Since ω, ω' are arbitrary, all the off-diagonal entries are the same. $\blacksquare$

Now we show that our selection mapping φ is unitary-invariant.

Lemma 6 For any $P \in \mathcal{P}(S)$ and $\Pi \in \mathcal{T}(S)$, $\varphi(\Pi P) = \Pi\varphi(P)$.

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

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

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

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

By homogeneity with respect to x, without loss of generality, 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

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

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 Unitary Transformations.

We similarly obtain a contradiction for the case $\varphi(\Pi P) \nsupseteq \Pi\varphi(P)$. $\blacksquare$

Below we translate the unitary invariance property to the Euclidian space of dimension $|S| - 1$, in steps. Recall the notation $e = (\frac{1}{|S|}, \cdots, \frac{1}{|S|})$. Define $\varphi^* : \mathcal{P}(S) - \{e\} \to \mathcal{P}(S) - \{e\}$ by

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

Lemma 7 For any $K \in \mathcal{P}(S) - \{e\}$ and $\lambda \geq 0$ with $\lambda K \in \mathcal{P}(S) - \{e\}$, $\varphi^*(\lambda K) = \lambda\varphi^*(K)$.

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 $|S| - 1$ dimensional linear subspace of $\mathbb{R}^S$ 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 $\mathcal{K}_e$. Also it preserves continuity in the Hausdorff metric. The claim below shows that φ^* satisfies additivity and translation invariance.

Lemma 8 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 $\lambda K, \lambda K' \in \mathcal{P}(S) - \{e\}$. By homogeneity shown in the previous lemma,

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

Then, we have

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

where the second line follows from mixture linearity of φ. ∎

For later use, we show the lemma below.

Lemma 9 Let $F : \Delta(S) \to \Delta(S)$ be a mixture-linear mapping satisfying $F(e) = e$. Then there is a unique bistochastic matrix Π such that $F(p) = \Pi p$ for every $p \in \Delta(S)$.

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 $j \in S$. 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 $i, j \in S$, this leads to $F_i(\delta_j) < 0$, which is a contradiction.

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

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

Lemma 10 Let $G : H_e \to H_e$ be a linear transformation such that $G(\Delta(S) - \{e\}) \subset \Delta(S) - \{e\}$ and there exists $\lambda_G \in (0, 1]$ such that $\|G(x)\| = \lambda_G \|x\|$ for any $x \in H_e$. Then, $\varphi^*(GK) = G\varphi^*(K)$ for all $K \in \mathcal{K}_e$.

Proof. Given G, define $F_G : \Delta(S) \to \Delta(S)$ by

$$F_G(p) = G(p - e) + e.$$

Then, it is easy to see that F_G takes values in $\Delta(S)$ and is mixture linear and $F_G(e) = e$. By Lemma 9, it has a representation by a doubly stochastic matrix Π_G and $F_G(p) = \Pi(p)$. Since F_G satisfies the unitary property, Π_G is in $\mathcal{T}(S)$.

By homogeneity of φ^*, without loss of generality we can take $K \in \mathcal{P}(S) - \{e\}$. By Lemma 9, G has a corresponding unitary transformation Π_G and $G(x) = \Pi_G(x + e) - e$ for any $x \in \Delta(S) - \{e\}$.

Then, from the unitary invariance property of φ we have

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

∎

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 $I \in \mathcal{I}$, one can take $\lambda > 0$ so that λI satisfies the assumption of Lemma 10. Conversely, any isometry is obtained from a matrix satisfying the assumption of Lemma 10.

By homogeneity of φ^*, we obtain

Lemma 11 The mapping φ^* is equivariant in isometries. That is, for any isometry $I \in \mathcal{I}$, $\varphi^*(I(K)) = I(\varphi^*(K))$.

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

Define $\varphi^{**} : \mathcal{K}^{|S|-1} \to \mathcal{K}^{|S|-1}$ by

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

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

Contraction representation

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

$$s^{**}(K) = \int_W \arg\max_{p \in K} \langle p, w \rangle \, \nu(dw)$$

where ν is the uniform distribution over W.[17]

Lemma 12 There exist $\varepsilon \geq 0$ and $\delta \geq 0$ such that

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

for every $K \in \mathcal{K}^{|S|-1}$.

Proof. Case 1 $|S| = 1, 2$: Obvious.

Case 2 $|S| = 3$: Since the image of a segment is its subsegment, we can apply Theorem 1.8 (b) in Schneider (1974) 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, (T_1 and T_2) must be the identity or the symmetry with respect to the origin respectively. Thus, without loss of generality

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

Case 3 $|S| \geq 4$: 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 (1974) so that we obtain

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

---

[17]Schneider (1974) has adopted a different definition of Steiner point, but it is equivalent to the current definition, which follows from Theorem 9.4.1 in Aubin and Frankowska (1990).

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

Finally, 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

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

indexed by $0 < \theta < \frac{\pi}{2}$. Then we have $s^{**}(K_\theta) = (\frac{\pi-\theta}{\pi}\sin\theta, 0, 0, \cdots, 0)$. Let $\overline{x}_1(K_\theta) = \max_{x \in \varphi^{**}(K_\theta)} x_1$, then we have $\overline{x}_1(K_\theta) = \frac{\pi-\theta}{\pi}\sin\theta + \varepsilon\frac{\theta}{\pi}\sin\theta + \delta\frac{\pi-\theta}{\pi}\sin\theta$. 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

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

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

$$\varphi(P) = \varepsilon\left[P - s(P)\right] + s(P).$$

Constancy of ε with regard to S

Let ε_S be the rate corresponding to S. When $S \subset S'$, since $P \in \mathcal{P}(S)$ implies $P \in \mathcal{P}(S')$, we must have $\varepsilon_S = \varepsilon_{S'}$. For every S, S', since $\varepsilon_S = \varepsilon_{S \cup S'}$, and $\varepsilon_{S'} = \varepsilon_{S \cup S'}$, we obtain the desired claim.

References

AHN, D. (2005): "Ambiguity without a state space," mimeo, University of California at Berkeley.

ANSCOMBE, F., AND R. AUMANN (1963): "A definition of subjective probability," *Annals of Mathematical Statistics*, 34, 199–205.

AUBIN, J.-P., AND H. FRANKOWSKA (1990): *Set-Valued Analysis*. Birkhauser, Boston.

BIRKOFF, G. (1946): "Three Observations on Linear Algebra," *Univ. Nat. Tucuman Ser A5*, pp. 147–151.

CHAMBERS, C., AND T. HAYASHI (2005): "Bayesian Consistent Prior Selection," Social Science WP 1238, California Institute of Technology.

CHASSAGNON, A., AND J.-C. VERGNAUD (1999): "A positive value of information for a non-bayesian decision-maker," in *Beliefs, interaction, and preferences*, ed. by M. Machina, and B. Munier. Kluwer.

CHATEAUNEUF, A., AND J. JAFFRAY (1989): "Some Characterizations of Lower Probabilities and Other Monotone Capacities through the Use of Möbius Inversion," *Mathematical Social Sciences*, 17, 263–283.

ELLSBERG, D. (1961): "Risk, ambiguity, and the Savage axioms," *Quarterly Journal of Economics*, 75, 643–669.

EPSTEIN, L. (1999): "A definition of uncertainty aversion," *Review of Economic Studies*, 66, 579–608.

EPSTEIN, L., AND T. WANG (1994): "Intertemporal asset pricing under Knightian uncertainty," *Econometrica*, 62(3), 283–322.

GAJDOS, T., J.-M. TALLON, AND J.-C. VERGNAUD (2004): "Decision making with imprecise probabilistic information," *Journal of Mathematical Economics*, 40(6), 647–681.

GHIRARDATO, P., AND M. MARINACCI (2002): "Ambiguity aversion made precise: a comparative foundation and some implications," *Journal of Economic Theory*, 102, 251–282.

GILBOA, I., AND D. SCHMEIDLER (1989): "Maxmin expected utility with a non-unique prior," *Journal of Mathematical Economics*, 18, 141–153.

GIRAUD, R. (2006): "Objective Imprecise Probabilistic Information, Second Order Beliefs and Ambiguity Aversion: an Axiomatization," mimeo, Université de Franche Comté.

JAFFRAY, J.-Y. (1989): "Linear utility for belief functions," *Operations Research Letters*, 8, 107–112.

KLIBANOFF, P., M. MARINACCI, AND S. MUKERJI (2005): "A Smooth Model of Decision Making Under Uncertainy," *Econometrica*, (6), 1849–1892.

KOPYLOV, I. (2006): "A parametric model of ambiguity hedging," mimeo, University of California at Irvine.

MACCHERONI, F., M. MARINACCI, AND A. RUSTICHINI (2006): "Ambiguity aversion, Robustness, and the Variational Representation of Preferences," *Econometrica*, 74, 1447–1498.

MUKERJI, S., AND J.-M. TALLON (2001): "Ambiguity aversion and incompleteness of financial markets," *Review of Economic Studies*, (68), 883–904.

OLSZEWSKI, W. (2007): "Preferences Over Sets of Lotteries," Discussion paper, Northwestern University, Forthcoming in *Review of Economic Studies*.

SCHNEIDER, R. (1974): "Equivariant Endomorphisms of the Space of Convex Bodies," *Transactions of the American Mathmatical Society*, 194, 53–78.

——— (1993): *Convex Bodies: the Brunn-Minkowski Theory*. Cambridge University Press, Cambridge, Mass.

STINCHCOMBE, M. (2003): "Choice with ambiguity as sets of probabilities," mimeo, University of Texas, Austin.

TAPKING, J. (2004): "Axioms for preferences revealing subjective uncertainty and uncertainty aversion," *Journal of Mathematical Economics*, 40, 771–797.

WANG, T. (2003): "A class of multi-prior preferences," Discussion paper, University British Columbia.