Center for
Mathematical Economics
Working Papers

480

May 2013

The von Neumann/Morgenstern approach to ambiguity

Martin Dumav and Maxwell B. Stinchcombe

Center for Mathematical Economics (IMW)
Bielefeld University
Universitätsstraße 25
D-33615 Bielefeld · Germany

e-mail: imw@uni-bielefeld.de
http://www.imw.uni-bielefeld.de/wp/
ISSN: 0931-6558

THE VON NEUMANN/MORGENSTERN APPROACH TO AMBIGUITY

MARTIN DUMAV, MAXWELL B. STINCHCOMBE

Abstract. A choice problem is risky (respectively ambiguous) if the decision maker is choosing between probability distributions (respectively sets of probability distributions) over utility relevant consequences. We provide an axiomatic foundation for and a representation of continuous linear preferences over sets of probabilities on consequences. The representation theory delivers: first and second order dominance for ambiguous problems; a utility interval based dominance relation that distinguishes between sources of uncertainty; a complete theory of updating convex sets of priors; a Bayesian theory of the value of ambiguous information structures; complete separations of attitudes toward risk and ambiguity; and new classes of preferences that allow decreasing relative ambiguity aversion and thereby rationalize recent challenges to many of the extant multiple prior models of ambiguity aversion. We also characterize a property of sets of priors, descriptive completeness, that resolves several open problems and allows multiple prior models to model as large a class of problems as the continuous linear preferences presented here.

> Roughly, risk refers to situations where the likelihood of relevant events can be represented by a probability measure, while ambiguity refers to situations where there is insufficient information available for the decision maker to assign probabilities to events. (Epstein and Zhang [22])

1. Introduction

This paper takes Epstein and Zhang's rough distinction as the defining difference between risky choice problems and ambiguous choice problems, and takes the "relevant events" to be sets of consequences. A risky decision problem is one in which the decision maker (DM) knows the probability distributions associated with their choices. An ambiguous decision problem is one in which the DM knows only partial descriptions of the probability distributions associated with their choices.

We identify a partial description of the probabilities with the *set* of probabilities satisfying the partial description. Under study are ambiguous decision problems in which the DM's preferences are continuous linear functionals on the class of compact sets of distributions over consequences.

1.1. **Comparison with Multiple Prior Models.** For the modeling of risky decisions, there are two main approaches: preferences over mappings from a state space to consequences, as in Savage [46]; or preferences over distributions over consequences, as in von Neumann and Morgenstern (vNM) [56]. The choice between the two is a question of convenience, but only if the prior is non-atomic. This follows from a change of variables and the result that for any non-atomic prior, p, and any distribution, μ, on a wide class of spaces, there is a measurable function such that $\mu = f(p)$. As to convenience, analyses of risky problems are essentially always taught and carried out in the space of distributions over consequences.

For multiple prior models of choice under ambiguity, descriptive completeness provides a condition analogous to a single prior being non-atomic — a set, S, of priors is descriptively complete if for any (relevant) set, A, of distributions over consequences, there is a measurable f

Date: May 31, 2013
Key words and phrases. Ambiguity, decision theory, multiple priors, descriptive completeness, continuous linear functionals on spaces of sets, constant and decreasing relative ambiguity aversion, zonoids.
Many thanks to Massimiliano Amarante, Svetlana Boyarchenko, Takashi Hayashi, Urmee Khan, Mark Machina, Marcin Peski, and Dale Stahl for help with this paper. They should be held blameless.

such that $A = f(S)$. Combining a descriptively complete set of priors with the same change of variables, one can model the same class of ambiguous problems either in the space of measurable functions or in the space of sets of distributions over consequences. The extant preferences over measurable functions, when expressed as preferences over sets of distributions, are either continuous and linear or are locally linear, leading to our focus on representing continuous linear functionals. The relative convenience of analyses in the space of distributions over consequences carries over to ambiguous problems, and the class of preferences studied here nests those studied in most of the multiple prior models.

1.2. **The Benefits of a Good Representation Theory.** As well as giving the set analogue of non-atomic priors, we give the representation theory for continuous linear preferences over sets of probabilities. The representation provides a number of results: the continuous linear preferences extend most of the previously studied multiple prior preferences; continuous linear preferences include those with decreasing (or increasing) relative ambiguity aversion, directly answering Machina's [38] challenges to many extant preferences-over-functions models of ambiguous choice; the form of the preferences generate new hypotheses about choice in the face of ambiguity; the representation yields characterizations of domains of problems on which α-minmax EU preferences are, and are not, ambiguity averse when $\alpha > \frac{1}{2}$; it allows for generalizations of first and second order stochastic dominance rankings to ambiguous decision problem; delivers a complete theory of updating convex sets of priors, and through this a Bayesian theory of the value of ambiguous information structures; complete separations between attitudes toward risk and attitudes toward ambiguity; and finally, with a representation theory for linear functionals in hand, we can begin the systematic study of the recently proposed preferences over sets of distributions that are non-linear but can be locally approximated by continuous linear functionals.

1.3. **Change of Variables in Risky Decision Problems.** Decision theory in the face of risk has two main models, related by change of variables. Both models use a space of consequences, $\mathbb{X}$, and one of them also has a measure space[1] of states, $(\Omega, \mathcal{F})$. In applications, $\mathbb{X}$ is often a compact subset of $\mathbb{R}$, and essentially always a Polish (complete separable metric) space. For this introductory section, we assume that $\mathbb{X}$ is compact as the more general case requires some details that impede acquiring an overview.

A preference ordering, $\succsim$, on $\Delta(\mathbb{X})$ is a complete transitive binary relation on $\Delta(\mathbb{X})$. von Neumann and Morgenstern (vNM) [56] gave a short axiomatic foundation for preferences over distributions on $\mathbb{X}$. Preferences satisfying their axioms have the property that $\mu \succsim \mu'$ iff

$$vNM(\mu) := \int_{\mathbb{X}} \boldsymbol{u}(x)\, d\mu(x) \geq vNM(\mu') := \int_{\mathbb{X}} \boldsymbol{u}(x)\, d\mu'(x) \tag{1}$$

where $\boldsymbol{u} \in C(\mathbb{X})$, the continuous functions on $\mathbb{X}$. Here, $\boldsymbol{u}$ is unique up to positive affine transformation.

By contrast, Savage's [46] work provides an axiomatic foundation for preferences over measurable functions from a state space, $(\Omega, \mathcal{F})$, to $\mathbb{X}$. The preferences over measurable functions $f, f' : \Omega \to \mathbb{X}$ can be represented by $f \succsim f'$ iff

$$Sav(f) := \int_{\Omega} \boldsymbol{u}(f(\omega))\, dp(\omega) \geq Sav(f') := \int_{\Omega} \boldsymbol{u}(f'(\omega))\, dp(\omega). \tag{2}$$

Here, the prior, p, a probability on $(\Omega, \mathcal{F})$, is uniquely determined, and $\boldsymbol{u}$ is, as before, unique up to positive affine transformation.

The approaches are directly related by change of variables, taking $\mu = f(p)$ (defined by $f(p)(E) = p(f^{-1}(E))$ for $E \subset \mathbb{X}$) and $\mu' = f'(p)$, the integrals on each side of the inequalities (1) and (2) are the same. The vNM approach specifies preferences over all of $\Delta(\mathbb{X})$, but, depending on the prior, p, this may or may not be true for Savage's approach. What is required for the class of models to be the same in (1) and (2) is that the prior, p, be **descriptively**

[1] A measure space is a non-empty set and a σ-field of subsets.

complete, that is, it is necessary that every $\mu \in \Delta(\mathbb{X})$ is of the form $f(p)$ for some measurable $f : \Omega \to \mathbb{X}$. For a single prior, descriptive completeness is equivalent to non-atomicity: the first part of Skorohod's representation theorem [52] implies that any non-atomic p is descriptively complete; if p has atoms and $\mathbb{X} = [0, M]$, then the set of $\mu \in \Delta(\mathbb{X})$ that are of the form $f(p)$ fails to contain a set with non-empty interior as well as a dense convex subset of $\Delta(\mathbb{X})$; and if $\mathbb{X}$ is finite and non-trivial, then the set fails to contain a subset of $\Delta(\mathbb{X})$ with non-empty interior.

1.4. **Change of Variables in Ambiguous Decision Problems.** The most widely used models of ambiguous decision problems involve a set, $S \subset \Delta(\Omega)$, of priors to describe preferences over measurable functions from Ω to $\mathbb{X}$. The same change of variables that relates (1) and (2) means that these preferences can be re-written as preferences over the **descriptive range** of S, denoted $\mathcal{R}(S)$ and defined as the class of $A \subset \Delta(\mathbb{X})$ that are of the form $f(S)$ for some $f : \Omega \to \mathbb{X}$. Such preferences are, mostly, the restrictions of continuous linear functions on the subsets of $\Delta(\mathbb{X})$ to $\mathcal{R}(S)$. We say that the set S is descriptively complete if its descriptive range is the class of all (relevant) subsets of $\Delta(\mathbb{X})$. If S is descriptively complete, models of preferences over functions from Ω to $\mathbb{X}$ and preferences over subsets of $\Delta(\mathbb{X})$ cover the same class of problems.

The first of the multiple prior models of preferences over functions $f : \Omega \to \mathbb{X}$ is due to Gilboa and Schmeidler [26]. Preferences satisfying their weakening of Savage's [46] axioms can be represented by $f \succsim f'$ iff

$$GS(f) := \min_{p \in S} \int_{\Omega} \boldsymbol{u}(f(\omega))\, dp(\omega) \geq GS(f') := \min_{p \in S} \int_{\Omega} \boldsymbol{u}(f'(\omega))\, dp(\omega) \tag{3}$$

for S a weakly closed, convex set of prior probabilities on Ω. If we let $A = f(S)$ and $B = f'(S)$, then change of variables (cov) delivers $A \succsim B$ iff

$$GS_{cov}(A) := \min_{\mu \in A} \int_{\mathbb{X}} \boldsymbol{u}(x)\, d\mu(x) \geq GS_{cov}(B) := \min_{\mu \in B} \int_{\mathbb{X}} \boldsymbol{u}(x)\, d\mu(x). \tag{4}$$

The function $A \mapsto GS_{cov}(A)$ is a continuous, linear functional on the class of closed subsets of $\Delta(\mathbb{X})$. For continuous linear preferences, every closed set is indifferent to its closed convex hull, so there is no loss in restricting attention to $\mathbb{K}_{\Delta(\mathbb{X})}$, the set of closed convex subsets of $\Delta(\mathbb{X})$. $GS_{cov}(\cdot)$ specifies preferences over all of $\mathbb{K}_{\Delta(\mathbb{X})}$, and the $GS(\cdot)$ preferences are the restriction of $GS_{cov}(\cdot)$ to the descriptive range of S, $\mathcal{R}(S)$.

There are three quite general subsequent versions of the multiple priors preferences. The first is the **$\boldsymbol{\alpha}$-minmax EU preferences** Ghirardato, Maccheroni, and Marinacci (GMM) [25], represented by $f \succsim f'$ iff

$$\begin{aligned} \alpha MEU(f) := \alpha_f \cdot \min_{p \in S} \int_{\Omega} \boldsymbol{u}(f(\omega))\, dp(\omega) + (1 - \alpha_f) \cdot \max_{q \in S} \int_{\Omega} \boldsymbol{u}(f(\omega))\, dq(\omega) \geq \\ \alpha MEU(f') := \alpha_{f'} \cdot \min_{p \in S} \int_{\Omega} \boldsymbol{u}(f'(\omega))\, dp(\omega) + (1 - \alpha_{f'}) \cdot \max_{q \in S} \int_{\Omega} \boldsymbol{u}(f'(\omega))\, dq(\omega) \end{aligned} \tag{5}$$

where S is again a weakly closed, convex set of probabilities on Ω.

If $f \mapsto \alpha_f$ is constant, then setting $A = f(S)$ and $B = f'(S)$, the change of variables delivers Olszewski's [42] preferences, $A \succsim B$ iff

$$\begin{aligned} \alpha MEU_{cov}(A) := \alpha \cdot \min_{\mu \in A} \int_{\mathbb{X}} \boldsymbol{u}(x)\, d\mu(x) + (1 - \alpha) \cdot \max_{\nu \in A} \int_{\mathbb{X}} \boldsymbol{u}(x)\, d\nu(a) \geq \\ \alpha MEU_{cov}(B) := \alpha \cdot \min_{\mu \in B} \int_{\mathbb{X}} \boldsymbol{u}(x)\, d\mu(x) + (1 - \alpha) \cdot \max_{\nu \in B} \int_{\mathbb{X}} \boldsymbol{u}(x)\, d\nu(a). \end{aligned} \tag{6}$$

Once again, $A \mapsto \alpha MEU_{cov}(A)$ is a continuous linear functional on the closed subsets of $\Delta(\mathbb{X})$, so there is no loss in restricting attention to $\mathbb{K}_{\Delta(\mathbb{X})}$. The set of problems that can be modeled in (5) and (6) is the same if S is descriptively complete. Further, Proposition 1 shows that unless the α-MEU preferences violate state independence, descriptive completeness implies that $f \mapsto \alpha_f$ must be constant in GMM's axiomatization.

A direct generalization of the α-MEU preferences are the **Monotonic, Bernoullian, and Archimedean (MBA)** preferences of Cerreia-Vioglio *et al.* [14]. They are also representable by the formula in (5), what differs are the restrictions on how α_f depends on f: α_f must be equal to $\alpha_{f'}$ if $p \mapsto \int \boldsymbol{u}(f)\,dp$ is a positive affine function of $p \mapsto \int \boldsymbol{u}(f')\,dp$ on S for α-MEU preferences, while they must be equal if the two functions are equal for MBA preferences. Proposition 2 shows that for MBA preferences, $\alpha_f = \alpha_{f'}$ if the minimum and maximum values of $\int \boldsymbol{u}(f)\,dp$ and $\int \boldsymbol{u}(f')\,dp$ are the same on S.[2] If the mapping $f \mapsto \alpha_f$ is non-constant but well-behaved, Example 4 shows that the cov version of MBA preferences may be smooth, i.e. locally approximatable by the continuous linear preferences under study here.

These observations lead us to two topics: a representation theorem for continuous linear functionals on $\mathbb{K}_{\Delta(\mathbb{X})}$; and the structure of descriptively complete sets of priors.

1.5. **The Representation Theorem.** The Riesz representation theorem tells us that $\Delta(\mathbb{X})$, the domain of the vNM preferences, is a subset of the dual space of $C(\mathbb{X})$, the finite signed measures. The Hahn-Jordan decomposition of the dual space tells us that $\Delta(\mathbb{X})$ is a spanning subset. Therefore *all* weak* continuous linear functionals $U : \Delta(\mathbb{X}) \to \mathbb{R}$ have an integral representation, $U(\mu) = \int_{\mathbb{X}} \boldsymbol{u}(x)\,d\mu(x)$ for some $\boldsymbol{u} \in C(\mathbb{X})$. This result is crucial to the study of choice in the presence of risk: monotonicity of $\boldsymbol{u}$ is equivalent to the preferences respecting first order dominance; monotonicity and concavity of $\boldsymbol{u}$ is equivalent to the preferences respecting second order dominance; for decision problems with actions $a \in A$ and a realization, x, of a random variable, the necessarily convex upper envelope of the linear functions $\{\mu \mapsto \int \boldsymbol{u}(a,x)\,d\mu(x) : a \in A\}$, is at the center of Blackwell's development of the value of information [9], [10]; log supermodularity of $\boldsymbol{u}(a,x)$ in an action a and a realization of a random variable x is a key ingredient in monotone comparative statics results [6].

Continuous linear functions on $\mathbb{K}_{\Delta(\mathbb{X})}$ have an integral representation that reduces to the vNM representation, (1), for risky problems, and contains the change of variables version of the Gilboa-Schmeidler and the GMM preferences, (4) and (6), as special cases. Theorem 1 shows that, modulo an infinitesimal caveat, continuous linear preferences on $\mathbb{K}_{\Delta(\mathbb{X})}$ are given by $A \succsim B$ iff

$$U(A) := \int_{\mathbb{U}_0^1} \min_{\mu \in A} \langle u, \mu \rangle \, d\eta_{\min}(u) + \int_{\mathbb{U}_0^1} \max_{\nu \in A} \langle v, \nu \rangle \, d\eta_{\max}(v) \geq \tag{7}$$
$$U(B) := \int_{\mathbb{U}_0^1} \min_{\mu \in B} \langle u, \mu \rangle \, d\eta_{\min}(u) + \int_{\mathbb{U}_0^1} \max_{\nu \in B} \langle v, \nu \rangle \, d\eta_{\max}(v),$$

where: $\mathbb{U}_0^1$ is the set of continuous functions with $\min_{x \in \mathbb{X}} u(x) = 0$, $\max_{x \in \mathbb{X}} u(x) = 1$; $\eta_{\min}$ and $\eta_{\max}$ are non-negative, countably additive measures with $(\eta_{\min} + \eta_{\max})$ normalized to be a probability; and $\langle f, \mu \rangle := \int_{\mathbb{X}} f(x)\,d\mu(x)$. This nests the previously discussed preferences as follows.

a. If $A = \{\mu\}$ and $B = \{\mu'\}$ are singleton sets, as they would be for risky decision problems, then (7) reduces to $\mu \succsim \mu'$ iff $\int_{\mathbb{X}} \boldsymbol{u}(x)\,d\mu(x) \geq \int_{\mathbb{X}} \boldsymbol{u}(x)\,d\mu'(x)$ where $\boldsymbol{u}$ is the **resultant of** $(\eta_{\min} + \eta_{\max})$, i.e. $\boldsymbol{u}(x) = \int_{\mathbb{U}_0^1} u(x)\,d(\eta_{\min} + \eta_{\max})(u)$.
b. For general $A, B \in \mathbb{K}_{\Delta(\mathbb{X})}$, then taking $\eta_{\min}$ and $\eta_{\max}$ to be the scaled point masses on the function $\boldsymbol{u}$, $\eta_{\min} = \alpha\delta_{\boldsymbol{u}}$ and $\eta_{\max} = (1-\alpha)\delta_{\boldsymbol{u}}$, (7) delivers the GMM preferences (6).
c. Taking $\alpha = 1$ yields the Gilboa-Schmeidler preferences given in (4).

The integral representation has many consequences: Corollaries 1.1 and 1.2 use it to characterize respect for first and/or second order stochastic dominance in ambiguous choice problems in terms of the support sets for $\eta_{\min}$ and $\eta_{\max}$; Corollary 1.3 uses it and first order dominance to bound the utility effects of ambiguity; Corollary 1.4 uses the integral representation to give the basic ordering result for ambiguous information structures; §5, especially Proposition 3,

[2]Amarante [2] gives the most general class of preferences satisfying the GMM axioms, those representable by the Choquet integral with respect to a capacity ξ, $U(f) = \int_S \boldsymbol{u}(f(P))\,d\xi(P)$.

uses it to provide complete separations of attitudes toward risk and attitudes toward ambiguity in two broad classes of problems.

1.6. **Descriptively Complete Sets of Priors.** There are several reasons that favor the use of descriptively complete sets of priors in economic models, though, with the exceptions Klibanoff [34] and Epstein and Ji [20], the sets of priors that have been used in the literature fail to be descriptively complete. First, without descriptive completeness, there are often severe limits to the set of problems that can be modeled. Second, these limits substantively affect the analyses. Third, the focus on sets of prior models rather than on sets of distributions over consequences has impeded our understanding of many issues, most especially comparisons of and non-constancy/constancy of degrees of ambiguity aversion.

Modeling with a set of priors that is not descriptively complete means that one is modeling a decision maker who cannot conceive of many, perhaps most, partially described sets of probabilities. Example 1 shows that a decision maker modeled as having any instance of commonly used class, S, of multiple priors can only conceive of a negligible set of problems when there are finitely many outcomes. The substantive effect of this limitation in the two outcome case is that all monotonic preferences, whether ambiguity loving, ambiguity averse, or neither, have exactly the same implications for everything in the descriptive range of S. For three or more outcomes, the negligibility of $\mathcal{R}(S)$ has arguably worse consequences. The general result in this direction is Theorem 3, which shows that if e.g. $\mathbb{X} = [0, M]$ and S fails to be descriptively complete, then $\mathcal{R}(S)$ misses at least a dense subset of $\mathbb{K}_{\Delta(\mathbb{X})}$.

The inability to distinguish behavioral differences between ambiguity loving or ambiguity averse behavior on the descriptive range of a set of priors is an example of the observation that properties of axioms restricted to small domains can be very different than their properties on larger domains. A second example is provided by Proposition 1, which shows that if the descriptive range of the set of priors, S, is all of $\mathbb{K}_{\Delta(\mathbb{X})}$ in GMM's α-MEU setting, then the mapping $f \mapsto \alpha_f$ in their representation must be constant. A third example is provided by Proposition 2, which shows that, under the same descriptive completeness condition applied to the set S in the MBA variant of α-MEU preferences, the mapping $f \mapsto \alpha_f$ can only depend on the upper and lower bounds of expected utility under S.

A further lesson contained in Example 1 is that enlarging the set of priors can shrink its descriptive range. This counter-intuitive result provides part of the explanation of why focus on multiple priors models rather than on sets of distributions over consequences has impeded our understanding. Another example of this kind of difficulty is apparent in Epstein [21], which shows that convexity of a capacity, hence non-emptiness of its core, S, is neither necessary nor sufficient for preferences over random variables to be ambiguity averse. In general, trying to identify degrees of ambiguity aversion by studying properties of sets of priors has not proved very fruitful. However, if one works with a descriptively complete set of priors, then change of variables delivers the same functional forms for the preferences, but now they are applied to $\mathbb{K}_{\Delta(\mathbb{X})}$ or to subclasses of $\mathbb{K}_{\Delta(\mathbb{X})}$. This allows us to demonstrate one source of the difficulties: Proposition 5 gives a class of sets, denoted $\mathbb{K}^{sym}_{\Delta(\mathbb{X})}$, encompassing many of the extant analyses, and shows that α-MEU preferences are ambiguity averse relative to this class of sets provided $\alpha > \frac{1}{2}$; it also shows that α-MEU preferences with $\alpha < 1$ cannot be ambiguity averse relative to any class $\mathcal{A} \subset \mathbb{K}_{\Delta(\mathbb{X})}$ if $\mathcal{A}$ contains the triangular sets of distributions over consequences.

1.7. **Nonlinear Functionals.** Machina [37] introduced the study of smooth preferences over distributions. These are locally linear, which means that vNM preferences provide local approximations, and the properties of the linear approximations determine the properties of the smooth preferences.[3] Many of the recently studied preferences for choice under ambiguity are not representable by linear functionals on sets of probabilities, but are representable as locally linear functionals. For example, variational preferences (e.g. [5], [36], or the tutorial [47]), are concave on $\mathbb{K}_{\Delta(\mathbb{X})}$, hence locally linear at most points in their domain.

[3]See [15] for a further development of the local approximation approach to smooth utility functions on probabilities.

1.8. **Outline.** The next section covers most of the main results of the paper in the case that there are just two consequences, $\#\mathbb{X} = 2$: representation; first order dominance; separation of risk and ambiguity attitudes; descriptive completeness and incompleteness; and local linear approximations to non-linear preferences. The subsequent section covers the representation theorem for continuous linear preferences. This leads to a theory of first and second order stochastic dominance for ambiguous problems as well as a theory of the value of ambiguous information. Section 4 gives the sufficient, and up to inessential duplications, necessary, condition for a set of priors to be descriptively complete. We show that: descriptively complete sets of priors have lower envelopes that mimic any concave or any convex capacity on finite sub-fields, providing further evidence that convexity of a capacity is not particularly related to ambiguity aversion outside of the two consequence case; and, combined with state independence, descriptive completeness implies the constancy of the α in the α-MEU preferences. Section 5 investigates the classes of problems for which there are good representations of the decomposition of preferences into attitudes toward risk and attitudes toward ambiguity.

The penultimate section shows how the continuous linear preferences discussed here resolve several puzzles and provide new classes of preferences. The first puzzle is whether or not α-MEU preferences can be ambiguity averse. We give a broadly useful class of problems for which α-MEU preferences with $\alpha \in (\frac{1}{2}, 1]$ are ambiguity averse, as well as a large class of problems where ambiguity aversion is only present when $\alpha = 1$. The second topic is the constancy of relative ambiguity aversion that is built into multiple prior preferences over random variables that satisfy a rank-dependence axiom — for any vNM utility function for risky problems, we give an associated infinite dimensional class of linear-in-sets preferences with decreasing (or increasing) relative ambiguity aversion. The third puzzle is how to update convex sets of probabilities, and the theory developed here leads to a Bayesian theory of the value of ambiguous information. Linear-in-sets preferences with decreasing (or increasing) relative ambiguity aversion are new. Also new are the class of preferences we give that respect a novel dominance relation for ambiguous problems, one that can distinguish between sources of uncertainty. The last section summarizes and indicates future directions.

Throughout, we reserve "Theorem" for results about the class of vNM preferences as a whole, and "Proposition" for results about subclasses of the vNM preferences.

2. Two Consequences

Urn problems are a particularly clear and compelling way to explain the intuitions for preferences in the presence of ambiguity, and that is where we begin.

2.1. **Urns and Interval Sets of Probabilities.** An urn is known to contain 90 balls, 30 of which are known to be Red, each of the remaining 60 can be either Green or Blue. The DM is faced with the urn, the description just given, and two pairs of choice situations.

(1) Choices between single tickets:
   (a) The choice between the Red ticket or the Green ticket.
   (b) The choice between the Red ticket or the Blue ticket.
(2) Choices between pairs of tickets:
   (a) The choice of the R&B or the G&B pair.
   (b) The choice of the R&G or the B&G pair.

In each situation, after the DM makes her choice, one of the 90 balls will be picked at random. If the ball's color matches the color of (one of) the chosen ticket(s), the DM gets $\$1,000$, otherwise they get nothing, a two-point set of consequences. Modal preferences in experiments are

$$R \succ G \text{ and } R \succ B, \text{ as well as}$$

$$R\&B \prec G\&B \text{ and } R\&G \prec B\&G.$$

People with these preferences cannot be assigning probabilities to these events if they prefer higher probabilities of better outcomes, for, if they did we would have

$$Pr(R) > Pr(G) \text{ and } Pr(R) > Pr(B), \text{ as well as}$$

$$Pr(R) + Pr(B) < Pr(G) + Pr(B) \text{ and } Pr(R) + Pr(G) < Pr(B) + Pr(G).$$

The probability that the Red ticket wins is $\frac{1}{3}$. That is, the action "choose Red" is risky, with the known probability $\frac{1}{3}$. The actions "choose Blue" and "choose Green" are ambiguous, leading to the interval of probabilities $[0, \frac{2}{3}]$. Choosing the B&G pair is risky, $\frac{2}{3}$, choosing the other two pairs is ambiguous, $[\frac{1}{3}, 1]$. The preferences $R \succ G$ and $R \succ B$ correspond to $\{\frac{1}{3}\} \succ [0, \frac{2}{3}]$, while the preferences $R\&B \prec G\&B$ and $R\&G \prec B\&G$ correspond to $[\frac{1}{3}, 1] \prec \{\frac{2}{3}\}$. A summary of this Ellsberg paradox is that people prefer knowing a probability p determines the chance that they win to knowing that the probability belongs to an interval with p at its center.

2.2. **Representation and Dominance.** In this urn problem, $\mathbb{X} = \{0, 1\}$, and $\Delta(\mathbb{X}) \subset \mathbb{R}^{\{0,1\}}$ can be represented by $[0, 1]$ where $q \in [0, 1]$ corresponds to the probability of receiving the better outcome. Let $\mathbb{K}_{\Delta(\mathbb{X})}$ be the class of non-empty closed, convex subsets of the probabilities $[0, 1]$, that is, $\mathbb{K}_{\Delta(\mathbb{X})} = \{[a, b] : 0 \leq a \leq b \leq 1\}$. In this case, continuous linear functionals on the convex sets of probabilities must be of the form $U([a, b]) = u_1 a + u_2 b$, $u_1, u_2 \in \mathbb{R}$.

An interval $[a, b]$ first order stochastically dominates $[a', b']$ if every expected utility who likes \$1,000 better than \$0 prefers the worst probability in $[a, b]$ to the worst in $[a', b']$ and prefers the best probability in $[a, b]$ to the best in $[a', b']$. This is equivalent to $a \geq a'$ and $b \geq b'$. For the utility function $U(\cdot)$ to respect first order dominance,[4] we must have $u_1, u_2 \geq 0$. Non-triviality of the preferences requires at least one inequality strict, and we normalize with $u_1 + u_2 = 1$.

Restricted to singleton sets of probabilities, U is a vNM utility function on $\{0, 1\}$. Since intervals with no width correspond to risky choices, the normalization gives $U([p, p]) = u_1 p + u_2 p = p$, e.g. $U([0, 0]) = 0$ and $U([1, 1]) = 1$. From this, the vNM utility function on $\mathbb{X} = \{0, 1\}$ is $u(0) = 0$ and $u(1) = 1$, which leads to GMM's α-minmax EU preferences by setting $u_1 = \alpha$, $u_2 = (1 - \alpha)$, and re-writing as

$$U([a, b]) = \alpha \cdot \left(\min_{\mu \in [a,b]} \int_{\mathbb{X}} u(x)\, d\mu(x) \right) + (1 - \alpha) \cdot \left(\max_{\nu \in [a,b]} \int_{\mathbb{X}} u(x)\, d\nu(x) \right). \tag{8}$$

2.3. **Separation of Risk and Ambiguity Attitudes.** In GMM's α-minmax EU preferences, $\alpha > \frac{1}{2}$, that is, $u_1 > u_2$, is often interpreted as ambiguity aversion. A change of basis in $\mathbb{K}_{\Delta(\mathbb{X})}$ allows us to see why this should be true in the two consequence case.[5] Rewriting $[a, b]$ as $[p - r, p + r]$, where $p = (a + b)/2$ and $r = (b - a)/2$, yields $U([p - r, p + r]) = (u_1 + u_2)p - (u_1 - u_2)r$, conveniently re-written as $U([p - r, p + r]) = p - vr$ with $v = (u_1 - u_2)$. Having $\alpha > \frac{1}{2}$ corresponds to $v > 0$, that is, to disliking expansions of the set of probabilities $[p - r, p + r]$ about the center p, capturing the modal preferencess.

In the utility function $U([p - r, p + r]) = p - vr$, we see an example of a complete separation between the attitude toward risk and the attitude toward ambiguity. The v measures the the tradeoff between risk and ambiguity, and any v can be combined with the expected utility part of the functional. Further, v can be elicited by giving people a choice between risky and ambiguous urns.

2.4. **Descriptive (In)Completeness.** For most modeling of random variables, one can take the probability space to be the unit interval with the uniform distribution, λ. This is because every probability distribution, μ, on a wide class of spaces (including every complete separable metric space) is the image measure, $f_\mu(\lambda)$, for an appropriately chosen random variable, f_μ. This is a domain equivalence result, it means that one can study random phenomena by studying distributions or by studying random variables, the choice is a matter of convenience. Further,

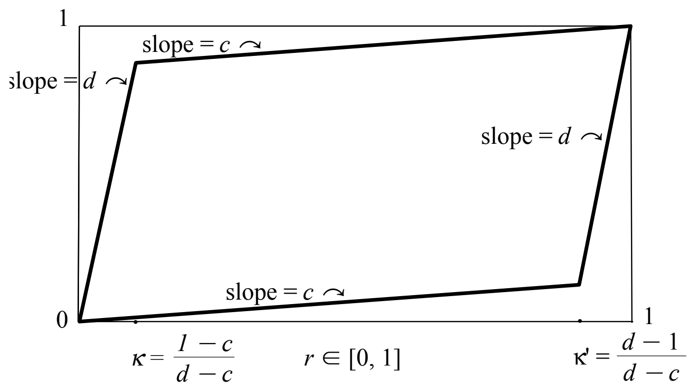
[4]With only two consequences, we cannot treat second order dominance here.

[5]Problems in which $\alpha > \frac{1}{2}$ can, and cannot, be regarded as ambiguity aversion for $\mathbb{X}$ having more than 2 elements are described in Proposition 5.

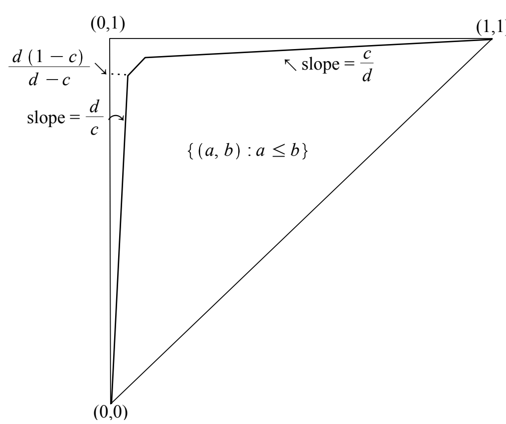
the unit interval can be replaced by any probability space that supports a countably additive non-atomic distribution.

More than non-atomicity is needed for modeling preferences under ambiguity.

Example 1. *Let* $(\Omega, \mathcal{F})$ *be* $([0,1], \mathcal{B})$*, the unit interval with the usual Borel* σ*-field, and let* λ *denote Lebesgue measure. Consider the set of priors* $S_{c,d} = \{p \in \Delta(\Omega) : c \leq dp/d\lambda \leq d\}$*,* $0 \leq c < 1 < d \leq \infty$*. Each set* $S_{c,d}$ *is weakly closed, convex, and has uncountably many linearly independent extreme points. Suppose the consequence space is* $\mathbb{X} = \{0, 1\}$*. The measurable functions* $f : \Omega \to \mathbb{X}$ *are of the form* $f(\omega) = 1_E(\omega)$*,* $E \in \mathcal{F}$*. Let* $\mathcal{R}(S_{c,d})$ *denote the set of* $A \in \mathbb{K}_{\Delta(\mathbb{X})}$ *that are of the form* $f(S_{c,d})$ *for some measurable* $f : \Omega \to \mathbb{X}$*. The possible lower bounds for the sets* $[a, b] \in \mathcal{R}(S_{c,d})$ *is given by the increasing, onto convex function* $\varphi(r) := \max\{cr, 1-d(1-r)\}$ *from* $[0,1]$ *to itself, and the upper bound is* $1-\varphi(1-r)$*, both given in Figure* 1*(a). Figure* 1*(b) gives the intervals* $[a, b] \in \mathcal{R}(S_{c,d})$ *as points in* $\{(a, b) : 0 \leq a \leq b \leq 1\}$*.*



Graph of $f(S_{c,d})$, $f = 1_E$, $\lambda(E) = r$



There are several lessons to be drawn from Example 1.

a. A preference ordering of the intervals $[a, b]$ that respects first order stochastic dominance must be increasing in both a and b. Restricted to the set $\mathcal{R}(S_{c,d})$ given in Figure 1(b), any increasing $(a, b) \mapsto U(a, b)$, ambiguity averse, ambiguity loving, or neither, and many non-monotonic $(a, b) \mapsto U(a, b)$ give the same ordering. The class of problems that a decision maker with priors $S_{c,d}$ can conceive of is so small that very different preferences have the same implications.
b. For $0 < c < c'$ and/or $d > d' > 1$, $S_{c,d}$ is a strict superset of $S_{c',d'}$, but the sets $\mathcal{R}(S_{c,d})$ and $\mathcal{R}(S_{c',d'})$ have only two points in common, certainty about the worst outcome, $[0, 0]$, and certainty about the best outcome, $[1, 1]$. Comparing attitudes toward ambiguity for two decision makers by comparison of the sets $S_{c,d}$ and $S_{c',d'}$ is not possible in any meaningful sense because the only risky problems decision makers have in common are those involving certainty of the outcome.
c. The class of problems that a decision maker can conceive of can disappear discontinuously. For example, as $c \downarrow 0$ and $d \uparrow \infty$, each $\mathcal{R}(S_{c,d})$ is uncountable, but in the limit, $\mathcal{R}(S_{0,\infty})$ is the three point set $\{[0, 0], [0, 1], [1, 1]\} \subset \mathbb{K}_{\Delta(\mathbb{X})}$.
d. The class of problems a decision maker can contemplate can be larger for a smaller set of priors. Consider the one-dimensional set of priors $S = \{p_\theta : \theta \in [0, 2]\} \subset S_{0,2}$ where each $p_\theta \in \Delta([0, 1])$ has a density with respect to Lebesgue measure equal to $2 - \theta$ on $[0, \frac{1}{2}]$ and equal to θ on $(\frac{1}{2}, 1]$. To see that $\mathcal{R}(S) = \mathbb{K}_{\Delta(\mathbb{X})}$, note that for any $A = [a, b] \in \mathbb{K}_{\Delta(\mathbb{X})}$, we can take $f_{a,b}(\omega) = 1_{[0,a/2)}(\omega) + 1_{[\frac{1}{2},\frac{1}{2}+b/2)}(\omega)$ so that $f_{a,b}(p_0)$ puts mass a on the better outcome, 1, while $f_{a,b}(p_2)$ puts mass b on the better outcome.
e. Looking briefly ahead to problems involving more than two consequences, let $\varphi : [0, 1] \to [0, 1]$ be increasing, onto, and convex. For a non-atomic P, define the capacity C by $C(E) = \varphi(P(E))$, and let $\Pi = \{p : (\forall E \in \mathcal{F})[p(E) \geq C(E)]\}$ be the core of C.

   Given the convexity of φ, the rank-dependent expected utility (RDU) of, equivalently, the Choquet expected utility of, a measurable $f : \Omega \to \mathbb{X}$ is

$$RDU(f) = \min_{p \in \Pi} \int_\Omega u(f(\omega))\, dp(\omega).$$

   The set of ambiguous outcomes that the decision maker can conceive of is the range set, $\mathcal{R}(\Pi) = \{[\varphi(r), 1 - \varphi(1 - r)] : r \in [0, 1]\}$, a negligible subset of the problems modeled using $\mathbb{K}_{\Delta(\mathbb{X})}$. As φ is onto, all minima are available in $\mathcal{R}(\Pi)$ when $\#X = 2$. Because RDU preferences only refer to the minimal utility, one might hope that the negligibility of the set of problems that the decision maker can consider is not problematic.

   However, when $\#\mathbb{X} \geq 3$, $\mathcal{R}(\Pi)$ will not, in any generality, contain the minima. To see why, let $x_1, x_2, x_3 \in \mathbb{X}$ with $x_1 \prec x_2 \prec x_3$, let $A \in \mathbb{K}_{\Delta(\mathbb{X})}$ be the set $\{\mu \in \Delta(\{x_1, x_2, x_3\}) : \mu(x_1) = \mu(x_2), \mu(x_3 \geq \frac{1}{3}\}$. A monotonic u_r having minimum 0 and maximum 1 is of the form $u_r = (u_r(x_1), u_r(x_2), u_r(x_3)) = (0, r, 1)$ for some $r \in [0, 1]$. For any such u_r, $\min_{\mu \in A} \langle u_r, \mu \rangle = \frac{1}{3} r + \frac{1}{3}$. If φ is moderately convex, then for e.g. $r = \frac{1}{2}$, there exists no $f : \Omega \to \{x_1, x_2, x_3\}$ such that $RDU(f) = \frac{1}{3} r + \frac{1}{3}$.

   For this kind of Choquet expected utility model to apply to different problems with more than two consequences, it may be necessary to choose a different function φ, and through it, a different set of priors for each decision problem. Such adjustment of preferences for different problems makes it difficult to convincingly examine how changes in circumstances change decisions [53].

2.5. **Nonlinearities.** There are interesting nonlinear preferences over the class of closed, not necessarily convex subsets of $\Delta(\mathbb{X})$, and interesting (quasi-)concave and strictly quasi-concave preferences over the class of closed convex subsets of $\Delta(\mathbb{X})$.

2.5.1. *All closed sets.* Continuity and linearity of preferences means that there is no loss in restricting preferences to the closed convex subsets of $\Delta(\{0, 1\})$, and this argument generalizes to more general spaces of consequences.

- Continuity means that if the (Hausdorff) distance between two sets is 0, then they are indifferent, and the distance between a set and its closure is 0.

- Linearity means that, taking A to be any closed subset of $\Delta(\{0,1\})$, $U(\frac{1}{2}A+\frac{1}{2}A) = U(A) = U(\sum_{i\le n}\frac{1}{n}A)$. Since $\sum_{i\le n}\frac{1}{n}A \to \mathrm{co}(A)$ where $\mathrm{co}(A)$ is the convex hull of A, we have $U(A) = U(\mathrm{co}(A))$.
- Strict concavity or quasi-concavity of the preferences over all subsets of $\Delta(\{0,1\})$ would lead to $\mathrm{co}(A) \succ A$ when A is not convex.

2.5.2. *Quasi-Concavity and Decreasing Ambiguity Aversion.* The intervals $[a,b]$ can be represented as points in the triangle $\{(a,b) \in \mathbb{R}^2 : 0 \le a \le b \le 1\}$ as in Figure 1(b). The linear utility functions $U([a,b]) = u_1 a + u_2 b$ have parallel, straight line indifference curves, and represent ambiguity averse perference that respect first order dominance iff $u_1 \ge u_2 \ge 0$. Geometrically, this corresponds to the slopes of the indifference curves belonging to $[-\infty,-1]$, with steeper/shallower curves corresponding to higher/lower degrees of ambiguity aversion.

Example 2 (Decreasing ambiguity aversion)**.** *In a fashion similar to Dekel's "fanning out" preferences* [17] *on* $\Delta(\mathbb{X})$*, specify monotonic preferences with non-parallel, straight-line indifference curves by joining the degenerate intervals* $[p,p]$ *to a point* $\mathbf{c}^\circ = (-x, 2+y)$ *for* $y > x$ *(see Figure* 2*). This yields a two-parameter class of preferences that are quasi-concave, not concavifiable, and which demonstrate decreasing ambiguity aversion as the choice set approaches certainty of the best outcome.*

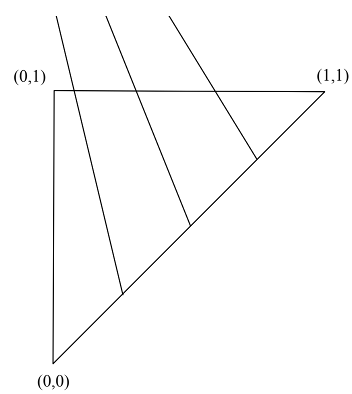


FIGURE 2. Decreasing ambiguity aversion

It is worth emphasizing that the decreasing ambiguity aversion preferences just specified are not linear-in-sets. With $\mathbb{X} = [0, M]$, we give linear-in-sets preferences with ambiguity aversion that is decreasing in wealth in §6.2.2.

2.5.3. *Concavity: Variational and MBA Preferences.* Preferences over $\mathbb{K}_{\Delta(\mathbb{X})}$ are quasi-concave if $[A \sim B] \Rightarrow [\alpha A + (1-\alpha)B] \succsim A$. Provided the set of priors is descriptively complete so that the domain of the preferences is the convex set $\mathbb{K}_{\Delta(\mathbb{X})}$, variational preferences ([5], [36], [47]) and Monotonic, Bernoullian, Archimedean (MBA, [14]) preferences are often concave, hence quasi-concave.

Example 3 (Variational)**.** *Variational preferences on* $\mathbb{K}_{\Delta(\mathbb{X})}$ *can be represented by* $V(A) = \min_{\mu\in A}\int_{\mathbb{X}} \boldsymbol{u}(x)\, d\mu(x) + c_A(\mu)$ *where each* c_A *as a convex function mapping* $\Delta(\mathbb{X})$ *to* $\mathbb{R}_+$*. A tractable parametrized version of these preferences for the two consequence case is given by* $V([p-r,p+r]) = \min_{q\in[p-r,p+r]}\{q + \frac{1}{\theta r}(p-q)^\alpha\}$ *for* $\alpha \in (1,2)$ *and* $\theta > 0$*. Solving and evaluating yields* $V([p-r,p+r]) = p - \kappa r^{\frac{1}{\alpha-1}}$ *for a parameter* $\kappa > 0$*. If* $\alpha \in (1,2)$*, then* $V(\cdot)$ *is smooth and concave, the linear approximations at* $[p-r,p+r]$ *are ambiguity neutral at* $r = 0$*, and become more ambiguity averse as* $r\uparrow$*.*

Proposition 2 shows that what matters for MBA preferences are the worst and the best expected utility in a set.

Example 4 (MBA)**.** *In the two outcome case, MBA preferences are given by* $MBA([a,b]) = \alpha_{a,b}a + (1-\alpha_{a,b})b$ *with* $\alpha_{a,a} \equiv \frac{1}{2}$*. A tractable parametrized version of these preferences for the two outcome case is given by* $\alpha_{a,b} = \frac{1}{2} + \kappa(b-a)^2$ *so that* $MBA([a,b]) = \frac{1}{2}(a+b) - \kappa(b-a)^3$*, which is strictly concave on* $\mathbb{K}_{\Delta(\mathbb{X})} = \{[a,b] : 0 \le a \le b \le 1\}$ *if* $\kappa > 0$*, strictly convex if* $\kappa < 0$*, and both linear and ambiguity neutral if* $\kappa = 0$*. At any* $[a,b]$*, the slope of the indifference curve is* $-(\frac{1}{2} + 3\kappa(b-a)^2)/(\frac{1}{2} - 3\kappa(b-a)^2)$*. This indicates that the preferences are ambiguity neutral in a neigborhood of the risky problems where* $a = b$*, and that, for* $\kappa > 0$*, the preferences become more ambiguity averse as* $r = (b-a)/2 \uparrow$*.*

The pattern of using information from the expected utility functionals that are the tangent approximations to non-linear preferences over $\Delta(\mathbb{X})$ comes from Machina [37]. In the same fashion, the vNM preferences studied in this paper are the tangent approximations to (almost everywhere) smooth nonlinear preferences. As such, properties of the non-linear preferences are often inherited from their linear approximations (see also [15]).

3. Representation Theory

We assume that the space of consequences, denoted $\mathbb{X}$, is Polish, that is, that it is a separable metric space with a topology that can be given by a metric making it complete. This section gives the representation theory for continuous linear preferences on the class of compact convex subsets of $\Delta(\mathbb{X})$, the set of countably additive Borel probabilities on (the Borel σ-field of subsets of) $\mathbb{X}$.

$C_b(\mathbb{X})$ denotes the set of bounded, continuous functions on $\mathbb{X}$ with the supnorm topology, and $\Delta(\mathbb{X})$ is a weak* closed and separable, convex subset of the dual space of $C_b(\mathbb{X})$. There are many metrics, e.g. the Prokhorov metric, that make $\Delta(\mathbb{X})$ complete, so that $\Delta(\mathbb{X})$ is also Polish. Because they induce the weak* topology, they have the property that $\mu_n \to \mu$ iff $\int u\, d\mu_n \to \int u\, d\mu$ for all $u \in C_b(\mathbb{X})$. $\mathbb{K}_{\Delta(\mathbb{X})}$ denotes the set of non-empty, compact, convex subsets of $\Delta(\mathbb{X})$ with the Hausdorff metric. It is well-known that $\mathbb{K}_{\Delta(\mathbb{X})}$ is compact (Polish) iff $\Delta(\mathbb{X})$ is compact (Polish) iff $\mathbb{X}$ is compact (Polish).

Let $\mathbb{U}_0^1 \subset C_b(\mathbb{X})$ denote the set $\{f \in C_b(\mathbb{X}) : \inf_{x\in\mathbb{X}} f(x) = 0, \sup_{x\in\mathbb{X}} f(x) = 1\}$. $\mathbb{X}$ is finite iff $\mathbb{U}_0^1$ is compact, and $\mathbb{X}$ is compact iff $\mathbb{U}_0^1$ is separable. Further, every $g \in C_b(\mathbb{X})$ has a unique representation of the form $rh + s$ for some $h \in \mathbb{U}_0^1$, $r \ge 0$, and $s \in \mathbb{R}$. $\mathcal{M}^s = \mathcal{M}^s(\mathbb{U}_0^1)$ denotes the set of countably additive, finite, signed measures on $\mathbb{U}_0^1$ with a separable support.

We identify a partial description of the probability distribution over $\mathbb{X}$ associated with a choice by the decision maker with the set of $\mu \in \Delta(\mathbb{X})$ that satisfy the partial description.

a. Because we study continuous linear preferences, there is no loss in assuming that each set is closed and convex.
b. When $\mathbb{X}$ is not compact, the assumption that the partially described sets are compact does entail a loss of generality.
c. By the Hahn-Banach theorem (in its supporting hyperplane form), a closed convex set, A, of probabilities can always be described as the set of μ given someone with expected utility function u at least utility $r_u^A := \min_{\nu\in A}\langle u, \nu\rangle$ where we let u range across $\mathbb{U}_0^1$.
d. For $A, B \in \mathbb{K}_{\Delta(\mathbb{X})}$ and $\beta \in (0,1)$, the partial description corresponding to the set $\beta A + (1-\beta)B$ is the set of μ given each expected utility maximizer $u \in \mathbb{U}_0^1$ at least utility $\beta r_u^A + (1-\beta)r_u^B$.
e. If both A and B can be defined using only finitely many $u \in \mathbb{U}_0^1$, then the same is true for $\beta A + (1-\beta)B$.

3.1. **Representation of Preferences.**

A weak* continuous rational preference relation on $\mathbb{K}_{\Delta(\mathbb{X})}$ is a complete, transitive relation, $\succsim$, such that for all $B \in \mathbb{K}_{\Delta(\mathbb{X})}$, the sets $\{A : A \succ B\}$ and $\{A : B \succ A\}$ are open. We will always assume that preferences on $\mathbb{K}_{\Delta(\mathbb{X})}$ are continuous and non-trivial. The continuous linear preferences are the ones that satisfy the following.

Axiom 1 (Independence)**.** *For all* $A, B, C \in \mathbb{K}_{\Delta(\mathbb{X})}$ *and all* $\beta \in (0,1)$*,* $A \succsim B$ *if and only if* $\beta A + (1-\beta)C \succsim \beta B + (1-\beta)C$*.*

An easy adaptation of standard arguments shows that a continuous rational preference relation on $\mathbb{K}_{\Delta(\mathbb{X})}$ satisfies the Axiom 1 if and only if it can be represented by a continuous linear functional. For $\eta^\circ \in \mathcal{M}^s(\mathbb{U}_0^1)$ and $A \in \mathbb{K}_{\Delta(\mathbb{X})}$, define $L_{\eta^\circ}(A) = \int_{\mathbb{U}_0^1} \min_{\mu \in A}\langle u, \mu\rangle \, d\eta^\circ(u)$, let $\mathcal{L}^\circ = \{L_{\eta^\circ} : \eta^\circ \in \mathcal{M}^s(\mathbb{U}_0^1)\}$, and let $\mathcal{L}$ be the completion of $\mathcal{L}^\circ$ in the metric $d(L, L') = \sup_{A \in \mathbb{K}_{\Delta(\mathbb{X})}} |L(A) - L'(A)|$.

Theorem 1. *$L : \mathbb{K}_{\Delta(\mathbb{X})} \to \mathbb{R}$ is continuous and linear if and only if $L \in \mathcal{L}$.*

Comments.

a. If $\mathbb{X}$ is finite or countable and discrete, then $\mathcal{L} = \mathcal{L}^\circ$, but if e.g. $\mathbb{X} = [0, M]$, then from [27, Theorem 3.11] one can show that $\mathcal{L} \setminus \mathcal{L}^\circ \neq \emptyset$. However, for the purposes of analyzing properties expressed using weak inequalities, such as ambiguity aversion or first/second order dominance, it is sufficient to analyze the dense subset, $\mathcal{L}^\circ$.

b. The measure η° in L_{η° has a Hahn-Jordan decomposition $\eta^\circ = \eta_+ - \eta_-$ where η_+ and η_- are non-negative measures with disjoint carriers, E_+ and E_-. Denote η_+ by $\eta_{\min}$, and let $\eta_{\max}$ denote the image of η_- under the mapping $f \mapsto (1 - f)$ from $\mathbb{U}_0^1$ to itself. Since $\max_{\mu \in A}\langle \mu, g\rangle = -\min_{\mu \in A}\langle \mu, -g\rangle$ for all $g \in \mathbb{U}_0^1$, up to the constant $\|\eta_-\|$, any L_{η° can be written as

$$L_\eta(A) = \int_{\mathbb{U}_0^1} \min_{\mu \in A}\langle u, \mu\rangle \, d\eta_{\min}(u) + \int_{\mathbb{U}_0^1} \max_{\nu \in A}\langle v, \nu\rangle \, d\eta_{\max}(v). \tag{9}$$

The carriers of $\eta_{\min}$ and $\eta_{\max}$ may overlap because $E_+ \cap (1 - E_-) \neq \emptyset$ is possible, but $E_{\min} \cap (1 - E_{\max})$ must be empty for carriers $E_{\min}$ and $E_{\max}$ of $\eta_{\min}$ and $\eta_{\max}$. This means that, subject to the carrier restriction just given, every element of $\mathcal{L}^\circ$ is, up to a positive affine transformation, of the form given in (9).

c. As $\eta_{\min}$ and $\eta_{\max}$ are both non-negative and at least one of them is non-null when L is non-trivial, the normalization $(\eta_{\min} + \eta_{\max})(\mathbb{U}_0^1) = 1$ is harmless, agrees with the $u_1 + u_2 = 1$ normalization in the two outcome case, and is maintained from now on. For notational simplicity, we will often denote a pair $(\eta_{\min}, \eta_{\max})$ by η.

d. Every $L \in \mathcal{L}^\circ$ has a resultant given by $\boldsymbol{u}(x) = \int_{\mathbb{U}_0^1} u(x) \, d(\eta_{\min} + \eta_{\max})(u)$. Every $L \in \mathcal{L}$ has a resultant because, restricted to the closed convex space of singleton sets, L is continuous and linear, and the Riesz representation theorem guarantees the existence of a resultant.

e. If the set A is replaced by a larger, more ambiguous one in (9), then the minimum term decreases and the maximum term increases. This suggests that the 'pessimistic' part of the preferences, $\eta_{\min}$, being larger than the 'optimistic' part, $\eta_{\max}$, should correspond to ambiguity aversion, that is, to a dislike of expansions of the sets of probabilities around its center. This is partially true.
   (i) Proposition 4 shows that $\eta_{\min}(E) \geq \eta_{\max}(E)$ for all E implies ambiguity aversion for the class of problems with decisions leading to centrally symmetric sets of probabilities.
   (ii) If $(\eta_{\min}, \eta_{\max}) = (\alpha \cdot \delta_{\boldsymbol{u}}, (1 - \alpha) \cdot \delta_{\boldsymbol{u}})$, Proposition 5 shows that $\eta_{\max} = 0$, i.e. $\alpha = 1$, is necessary for the preferences to be ambiguity averse for any class of problems that includes the triangular sets of probability distributions.
   (iii) $\eta_{\min}$ and $\eta_{\max}$ having different support sets can mean that the preferences are not ambiguity averse relative to the simplest class of ambiguous problems. For example, if $\eta = (\alpha \cdot \delta_u, (1 - \alpha) \cdot \delta_v)$ for $u \neq v$ and $\alpha \in (0, 1)$, then L_η is neither ambiguity averse nor ambiguity loving relative to the class of line segments $A = [\![\mu, \nu]\!] = \{(1 - \alpha)\mu + \alpha\nu : \alpha \in [0, 1]\}$.

The argument for Theorem 1 when $\mathbb{X}$ is a finite is much easier and directly shows that $\mathcal{L} = \mathcal{L}^\circ$ in this case. The more involved proof for the general Polish case is in the Appendix.

Proof of Theorem 1 *when* $\mathbb{X}$ *is finite.* The linearity of L_{η° is immediate, its continuity follows from the theorem of the maximum and dominated convergence.

Now suppose that $L : \mathbb{K}_{\Delta(\mathbb{X})} \to \mathbb{R}$ is continuous and linear. For each $A \in \mathbb{K}_{\Delta(\mathbb{X})}$ and f in the finite dimensional, compact set $\mathbb{U}_0^1$, define the support function $\psi_A(f) = \min_{\mu \in A}\langle f, \mu\rangle$. Each support function belongs to $C(\mathbb{U}_0^1)$, the set of continuous functions on $\mathbb{U}_0^1$ with the sup norm.

Because $d_H(A,B) = \sup\{|\psi_A(f) - \psi_B(f)| : \|f\|_\infty = 1\}$, $A \leftrightarrow \psi_A$ is an isometric isomorphism between $\mathbb{K}_{\Delta(\mathbb{X})}$ and the support functions. From Hörmander [28, Theorem 9], the span of the set of support function is a vector lattice of functions separating points in $\mathbb{U}_0^1$ to arbitrary values. Because $\mathbb{X}$ is finite, $\mathbb{U}_0^1$ is compact, and the Stone-Weierstrass theorem implies that the span is dense in $C(\mathbb{U}_0^1)$. Continuous functionals are determined by their values on dense subspaces. By the Riesz representation theorem, a weak*-continuous linear functional on $C(\mathbb{U}_0^1)$ has a unique representation as an integral against an $\eta^\circ \in \mathcal{M}^s(\mathbb{U}_0^1)$. □

3.2. **First and Second Order Dominance.** In expected utility analysis, the concepts of first and second order stochastic dominance play a central role. Theorem 1 allows us to extend these concepts to ambiguous choice problems. We let $\mathbb{ND}$ denote the set of non-decreasing functions in $\mathbb{U}_0^1$ and $\mathbb{NDC} \subset \mathbb{ND}$ the set of non-decreasing concave functions.

Definition 1. *For $\mu, \nu \in \Delta([0,M])$: μ **first order dominates** ν, written $\mu \succsim_F \nu$, if for all $u \in \mathbb{ND}$, $\langle u, \mu\rangle \geq \langle u, \nu\rangle$; and μ **second order dominates** ν, written $\mu \succsim_S \nu$, if for all $u \in \mathbb{NDC}$, $\langle u, \mu\rangle \geq \langle u, \nu\rangle$.*

Thus, $\mu \succsim_F \nu$ iff every expected utility maximizer with monotonic preferences over certain outcomes prefers μ to ν, and $\mu \succsim_S \nu$ if every risk-averse expected utility maximizer with monotonic concave preferences prefers μ to ν.

Definition 2. *For sets $A, B \in \mathbb{K}_{\Delta([0,M])}$, we say that A **first (resp. second) order dominates** B, written $A \succsim_F B$ (resp. $A \succsim_S B$), if for all $u \in \mathbb{ND}$ (resp. all $u \in \mathbb{NDC}$), $\min_{\mu_A \in A}\langle u, \mu_A\rangle \geq \min_{\mu_B \in B}\langle u, \mu_B\rangle$, and $\max_{\nu_A \in A}\langle u, \nu_A\rangle \geq \max_{\nu_B \in B}\langle u, \nu_B\rangle$.*

Associated with each $A \in \mathbb{K}_{\Delta(\mathbb{X})}$ are the two support functions, the concave $\psi_A(u) = \min_{\mu \in A}\langle u, \mu\rangle$ and the convex $\psi^A(u) = \max_{\nu \in A}\langle u, \nu\rangle$. In terms of these functions, $A \succsim_F B$ iff for all $u \in \mathbb{ND}$, $\psi_A(u) \geq \psi_B(u)$ and $\psi^A(u) \geq \psi^B(u)$. Further, $L_\eta \in \mathcal{L}^\circ$ respects first order dominance iff for all $A \succsim_F B$,

$$\int_{\mathbb{U}_0^1} (\psi_A - \psi_B)(u)\, d\eta_{\min}(u) \geq \int_{\mathbb{U}_0^1} (\psi^B - \psi^A)(u)\, d\eta_{\max}(u). \tag{10}$$

Further, the closure, in $\mathcal{L}$, of the set of L_η with η satisfying this condition, are exactly those that respect first order dominance. Unfortunately, (10) is somewhat difficult to work with: $(\psi_A - \psi_B)$ is the difference of concave functions, so may be concave, convex, or neither; $(\psi^B - \psi^A)$ is the difference of convex functions, so may concave, convex, or neither. Matters are simpler if $\eta_{\min}$ and $\eta_{\max}$ are carried by $\mathbb{ND}$.

Corollary 1.1. *$A \succsim_F B$ iff $L_\eta(A) \geq L_\eta(B)$ for all η with $(\eta_{\min} + \eta_{\max})(\mathbb{ND}) = 1$.*

Proof. Rearranging terms, $L_\eta(A) \geq L_\eta(B)$ iff

$$\int (\psi_A - \psi_B)(u)\, d\eta_{\min}(u) \geq \int (\psi^B - \psi^A)(u)\, d\eta_{\max}(u).$$

If $A \succsim_F B$, then for all $u \in \mathbb{ND}$, $(\psi_A - \psi_B)(u) \geq 0 \geq (\psi^B - \psi^A)(u)$.

Considering η's of the form $\eta_{\min} = \delta_u$ and $\eta_{\max} = \delta_u$, $u \in \mathbb{ND}$ gives the reverse implication. □

While $(\eta_{\min} + \eta_{\max})(\mathbb{ND}) = 1$ is sufficient for L_η to respect first order dominace, it is not necessary.

Example 5. *For $u, v \in \mathbb{ND}$, $u \neq v$, let $(\eta_{\min}, \eta_{\max}) = (\alpha\delta_u, (1-\alpha)\delta_{1-v})$. For any $A \succsim_F B$, we have $L_\eta(A) \geq L_\eta(B)$.*

For the following, we replace $\mathbb{ND}$ with $\mathbb{NDC}$ in Corallary 1.1.

Corollary 1.2. *$A \succsim_S B$ iff $L_\eta(A) \geq L_\eta(B)$ for all η with $(\eta_{\min} + \eta_{\max})(\mathbb{NDC}) = 1$.*

Note that Example 5 goes through if $u, v \in \mathbb{NDC}$ and $u \neq v$, showing that $(\eta_{\min} + \eta_{\max})(\mathbb{NDC}) = 1$ is sufficient, but not necessary for L_η to respect second order dominance. Working with $A \succsim_S B$ in (10) gives a characterization of the L_η that respect second order dominance.

3.3. **A Balance Interpretation of Respecting Dominance.** We work with $\mathbb{X} \subset \mathbb{R}$ and preferences satisfying the usual order. Working with different orders on more general spaces of consequences can be done with comonotonicity.

Example 6. *Suppose that* $\mathbb{X} = \{0, M\} \subset [0, M]$ *so that* $C_b(\mathbb{X}) = \{(f_0, f_M) : f_0, f_M \in \mathbb{R}\}$*,* $\mathbb{U}_0^1 = \{(0,1),(1,0)\}$*, and* $\mathbb{ND} = \{(0,1)\}$*. If L respects first order dominance, then* $\alpha := \eta_{\min}(0,1) \geq 0$ *and* $(1-\alpha) := \eta_{\max}(0,1) \geq 0$*, delivering* $L([a,b]) = \alpha \cdot a + (1-\alpha) \cdot b$ *as in the analysis of §2. The change of basis* $[a,b] = [p-r, p+r]$ *delivers* $L([a,b]) = p - vr$ *where* $v = 2\alpha - 1$ *and* $|v| \leq 1$*.*

An alternative interpretation of $|v| \leq 1$ in the two consequence case is that the preferences are **balanced** in the sense that for any interval $[a,b]$, $[a,a] \precsim [a,b] \precsim [b,b]$. When there are many consequences, $\mathbb{ND}$-unanimity about the best and the worst point in a set of distributions is not generally available, and this balance interpretation of first order dominance disappears. However, respecting first order dominance does constrain the relation between the range of the resultant and the range of L.

Corollary 1.3. *If* $L \in \mathcal{L}$ *respects first order stochastic dominance, then for any* $A \in \mathbb{K}_{\Delta([\underline{m},\overline{m}])}$*,* $\boldsymbol{u}(\underline{m}) \leq L(A) \leq \boldsymbol{u}(\overline{m})$ *where* $\boldsymbol{u}$ *is the resultant* L*.*

Proof. $\{\delta_{\underline{m}}\} \precsim_F A \precsim_F \{\delta_{\overline{m}}\}$, $L(\{\delta_{\underline{m}}\}) = \boldsymbol{u}(\underline{m})$, and $L(\{\delta_{\overline{m}}\}) = \boldsymbol{u}(\overline{m})$. □

3.4. **The Value of Ambiguous Information.** For an expected utility maximizing decision maker facing a risky problem the information they will have when making a decision can be encoded in a posterior distribution, $\beta \in \Delta(\mathbb{X})$. The value of β is $V_u(\beta) = \max_{a \in A} \int u(a,x)\, d\beta(x)$ where $u : A \times X \to \mathbb{R}$.

A prior is a point $p \in \Delta(\mathbb{X})$, and an information structure is a dilation of p, that is, a distribution, $Q \in \Delta(\Delta(\mathbb{X}))$, such that $\int \beta\, dQ(\beta) = p$. The value of the information structure is given by $V_u(Q) := \int_{\Delta(\mathbb{X})} V_u(\beta)\, dQ(\beta)$. An information structure Q dominates Q' if for all u, $V_u(Q) \geq V_u(Q')$, equivalently, if for all convex $V : \Delta(\mathbb{X}) \to \mathbb{R}$, $\int_{\Delta(\mathbb{X})} V(\beta)\, dQ(\beta) \geq \int_{\Delta(\mathbb{X})} V(\beta)\, dQ'(\beta)$.

For vNM utility maximizing decision maker facing an ambiguous problem, the information they will have when making a decision can be encoded in a set of posterior distributions, $B \in \mathbb{K}_{\Delta(\mathbb{X})}$. The value of B is $V_U(B) = \max_{a \in A} U(\delta_a \times B)$ where $U : A \times \mathbb{K}_{\Delta(\mathbb{X})} \to \mathbb{R}$ is a continuous linear functional on compact convex subsets of $\Delta(A \times \mathbb{X})$ of the form $\delta_a \times B$ (where δ_a is point mass on a).

A set-valued prior is a set $A \in \mathbb{K}_{\Delta(\mathbb{X})}$, and an information structure is a distribution, $Q \in \Delta(\mathbb{K}_{\Delta(\mathbb{X})})$, such that $\int_{\mathbb{K}_{\Delta(\mathbb{X})}} B\, dQ(B) = A$. It is very important to note the domain over which we integrate here, it is $\mathbb{K}_{\Delta(\mathbb{X})}$, not $\Delta(\mathbb{X})$.[6] The value of the information structure Q is given by $V_U(Q) := \int_{\mathbb{K}_{\Delta(\mathbb{X})}} V_U(B)\, dQ(B)$. As above, an information structure Q dominates Q' if for all U, $V_U(Q) \geq V_U(Q')$. The usual argument that convex functions are the upper envelope of the affine functions they majorize delivers the following.

Corollary 1.4. *Q dominates Q' iff* $\int v(B)\, dQ(B) \geq \int v(B)\, dQ'(B)$ *for every convex* $v : \mathbb{K}_{\Delta(\mathbb{X})} \to \mathbb{R}$*.*

Here we follow the standard Bayesian approach and model information structures as dilations. By contrast, previous work has limited the class of priors, A, and then studied a special class of dilations of each $p \in A$. The set of A for which this can be done is non-generic in both the measure theoretic and the topological sense, and the problems that one can consider are limited to ones in which the decision maker will learn only that the true value belong to some $E \subset \mathbb{X}$. Here, A is expressed as a convex combination of/integral of B's in $\mathbb{K}_{\Delta(\mathbb{X})}$, and this is what makes the problem tractable. Section 6.3 uses this insight to present a fairly complete solution to the problem of convincing a Bayesian having ambiguous information, a problem that cannot be sensibly modeled with the previous approaches.

[6]See e.g. [41, §2] for a quick development of the expectation of random closed sets.

4. Descriptively Complete Sets of Priors

In order for multiple prior models to cover as broad a range of choice situations as preferences over closed convex sets of probabilities, the set of priors should be descriptively complete.

Definition 3. *A set of probabilities,* Π*, on a measure space* $(\Omega, \mathcal{F})$ *is* ***descriptively complete*** *if for any Polish* $\mathbb{X}$ *and any* $A \in \mathbb{K}_{\Delta(\mathbb{X})}$*, there exists a measurable* $f_A : \Omega \to \mathbb{X}$ *such that* $\{f_A(p) : p \in \Pi\} = A$*.*

If Π is descriptively complete, then so is $\overline{\text{co}}(\Pi)$, the weak closure of its convex hull. A set of priors, Π, can be too large or too small to be descriptively complete. A **standard measure space** is one that is measurably isomorphic to a measurable subset of a Polish space. The Borel isomorphism theorem (e.g. [18, III.17]) tells us that all uncountable standard measure spaces are measurably isomorphic.

Example 7. *Suppose that* $(\Omega, \mathcal{F})$ *is an uncountable standard probability space. If* $\Pi = \Delta(\Omega)$*, then for any* $f : \Omega \to \mathbb{X}$*,* $f(\Pi) = \Delta(R_f)$ *where* R_f *is the range of* f*. If* $\Pi = \{\alpha p + (1-\alpha)p' : \alpha \in [0,1]\}$*, then for any* $f : \Omega \to \mathbb{X}$*,* $f(\Pi)$ *is either a* 0*- or a* 1*-dimensional subset of* $\Delta(\mathbb{X})$*.*

4.1. Measurable Identifiability.

Breiman et al. [13] show that the following condition is necessary and sufficient for the existence of consistent estimators.

Definition 4. *A measurable* $\Pi \subset \Delta_{\mathcal{F}}$ *is* ***measurably identifiable*** *if there exists an* $E \in \mathcal{F}$*, and a measurable, onto* $\varphi : E \to \Pi$ *such that for all* $p \in \Pi$*,* $p(\varphi^{-1}(p)) = 1$*.*

Measurable identifiability is a strengthened form of mutual orthogonality — for $p \neq q \in \Pi$, $p(\varphi^{-1}(p)) = 1$, $q(\varphi^{-1}(q)) = 1$, and $\varphi^{-1}(p) \cap \varphi^{-1}(q) = \varnothing$. The connection to consistent estimation can be seen as follows: let $\Pi = \{p_r : r \in [0,1]\}$ where $p_r \in \Delta(\{0,1\}^{\mathbb{N}})$ is the distribution of an i.i.d. sequence of Bernoulli(r) random variables; define $\varphi'(\omega) = \limsup_n \frac{1}{n}\#\{k \leq n : \omega_k = 1\}$; set $\varphi(\omega) = p_{\varphi'(\omega)}$; by the strong law of large numbers, $p_r(\varphi^{-1}(p_r)) = 1$ for each p_r; $\varphi'_n(\omega) := \frac{1}{n}\#\{k \leq n : \omega_k = 1\} \to \varphi'(\omega)$ with p_r mass 1; and $p_{\varphi'_n}$ is a consistent sequence of estimators.

The following minimalist example satisfying Definition 4 will appear several times below.

Example 8. *Let* $E = \Omega = [0,1] \times [0,1]$*, for each* $r \in [0,1]$*, takes* p_r *to be the uniform distribution on* $\{r\} \times [0,1]$*, take* $\Pi^\circ = \{p_r : r \in [0,1]\}$*, and set* $\varphi(r,u) = p_r$*. By the Borel isomorphism theorem, there exists* $\xi_A : [0,1] \leftrightarrow A$*, that is one-to-one, onto, measurable, with a measurable inverse. By the Blackwell and Dubins* [11] *extension of the Skorohod representation theorem, there exists a jointly measurable* $\boldsymbol{b} : \Delta(\mathbb{X}) \times [0,1] \to \mathbb{X}$ *with the property that for all* μ*,* $\boldsymbol{b}(\mu, \lambda) = \mu$ *where* λ *is the uniform distribution on* $[0,1]$*. Defining* $f_A(r,u) = \boldsymbol{b}(\xi_A(r), u)$ *delivers* $f_A(\Pi^\circ) = A$*.*

Theorem 2. *If* S *is an uncountable, measurably identifiable set of non-atomic priors on a standard space, then it is descriptively complete.*

Comments.

a. We will see that, up to inessential duplication, measurable identifiability is also necessary for descriptive completeness.
b. Measurably identifiable sets of non-atomic priors satisfy Siniscalchi's [51] characterization of plausible sets of priors.
c. An outline of the proof of Theorem 2 is contained in Example 8. Stronger versions of the result that include a continuity result can be found in [19, §2]. A discussion of how to dispense with the standardness assumption, at the cost of conditions that are slightly more complicated to state, can be found in [19, §3].

4.2. Properties of Descriptively Complete Sets.

We now give some basic properties of descriptively complete sets: Corollary 2.1 shows that the lower envelope of a descriptively complete set of probabilities is solvable in Wakker's [57] sense, and is never a convex capacity; this non-convexity is also a consequence of Corollary 2.2, which shows the lower envelope contains within it all convex capacities, and all concave capacities, on finite partitions of Ω; Corollary 2.3 shows that a descriptively complete set can not be expressed as a set of densities with respect to a σ-finite measure; and Corollary 2.4 shows that up to inessential duplications, measurable identifiability is nearly necessary for descriptive completeness.

4.2.1. *Lower Envelopes of Descriptively Complete Sets.* The lower envelope of a set of probabilities S is the capacity defined by $c_S(E) = \inf\{p(E) : p \in S\}$. Wakker [57] calls a capacity C **solvable** if for each $E \subset G$, $E, G \in \mathcal{F}$, and each $\gamma \in (C(E), C(G))$, there exists an $F \in \mathcal{F}$ such that $E \subset F \subset G$ and $C(F) = \gamma$.

Corollary 2.1. *If S is descriptively complete, then c_S is solvable, and not convex.*

Proof. If S is descriptively complete, then there exists $f : \Omega \to [0,1]^2$ such that $f(S) = T := \overline{\text{co}}(\Pi^\circ)$ where Π° was given in Example 8.

Solvability: Suppose first that there exists $p^* \in S$ such that $p^*(E) = c_S(E)$. Since $p^*(G) \geq c_S(G)$ and p^* is non-atomic, there is a subset, F^*, of the intersection of E with a carrier of p^*, such that $p^*(F^*) = \gamma$. Let $F = E \cup F^*$. If the infimum is not achieved, i.e. for no p^* is $p^*(E) = c_S(E)$, take a sequence p^*_n with $c_S(G) > p^*_n(E) \downarrow c_S(E)$, for each $n \in \mathbb{N}$, pick F^*_n as before and let $F = E \cup \bigcup_{n\in\mathbb{N}} F^*_n$.

Let $A' = \{(r,u) : u \geq \frac{1}{2}\}$, $B' = \{(r,u) : \frac{1}{2}r \leq u \leq \frac{1}{2} + \frac{1}{2}r\}$, and let $A = f^{-1}(A')$, $B = f^{-1}(B')$. It is immediate that $c_T(A') = c_T(A' \cup B') = \frac{1}{2}$, while $c_T(B') = \frac{1}{2} > c_T(A' \cap B') = 0$, which yields $c_T(A' \cup B') + c_T(A' \cap B') = \frac{1}{2} < c_T(A') + c_T(B') = 1$, showing that $c_T(\cdot)$ is not convex. Since each $q \in S$ has image measure $f(q) \in T$, $c_S(A \cup B) + c_S(A \cap B) = \frac{1}{2} < c_S(A) + c_S(B) = 1$. □

Corollary 2.2. *If S is descriptively complete, then for any convex (or concave) capacity C on the set of subsets of finite set, there exist a finite partition of Ω such that the restriction of c_S to the partition is isomorphic to C.*

Proof. We prove this for a convex capacity on the non-empty sets E, E^c. Induction completes the proof, and the argument for concave capacities is essentially identical.

If S is descriptively complete, then there exists $f : \Omega \to [0,1]^2$ such that $f(S) = \overline{\text{co}}(\Pi^\circ)$ where Π° was given in Example 8. Suppose that $C(E) = a$, $C(E^c) = b$, $a \leq b$, $a + b \leq 1$. If $A' = \{(r,u) : u \leq r(1-b) + (1-r)a\}$, then $c_T(A) = a$ and $c_T(A^c) = b$. □

Easy variants of this argument cover the mixed convex-concave (*cavex* in the terminology of [57]) capacities on finite partitions that appear in the α-MEU preferences over functions from a finite state space to consequences [29, 30, 31].

4.2.2. *Descriptively Complete Sets Are Not Dominated.* The descriptively complete set of priors in Example 8 has an uncountable set of extreme points, and the extreme points have disjoint supports. This yields the following.

Corollary 2.3. *No σ-finite measure can dominate a descriptively complete set.*

Proof. If S is descriptively complete, then there exists a measurable $f : \Omega \to [0,1]^2$ such that $f(S) = \overline{\text{co}}(\Pi^\circ)$ (where Π° was given in Example 8); for each $r \neq r'$, each q_r in $f^{-1}(p_r) \in S$ and each $q_{r'}$ in $f^{-1}(p_{r'}) \in S$ must be non-atomic, and the pair must be mutually orthogonal. No σ-finite measure can assign strictly positive mass to each of the uncountably many disjoint carrier sets. □

4.2.3. *Measurable Non-Identifiability and Descriptive Completeness.* Being measurably identifiable is sufficient for descriptive completeness. Sets of priors can fail to be measurably identifiable but still be descriptively complete because one can duplicate coverage, that is, one can send many priors to the same distribution over consequences.

Example 9. *Let $E = \Omega = [0,1] \times [0,1]$, for each $r \in [0, \frac{1}{2}] \cup \{1\}$, let p_r be the uniform distribution on $\{r\} \times [0,1]$, and let $\Pi^\dagger = \{p_r : r \in [0, \frac{1}{2})\} \cup \{\alpha p_{\frac{1}{2}} + (1-\alpha)p_1 : \alpha \in [0,1]\}$. For all $\alpha \in (0,1)$, the probabilities $\alpha p_{\frac{1}{2}} + (1-\alpha)p_1$ assign positive mass to all relatively open* *subsets of* $\{\frac{1}{2}, 1\} \times [0,1]$ *so* $\Pi^\dagger$ *cannot be measurably identifiable. However, for any* $A \in \mathbb{K}_{\Delta(\mathbb{X})}$*, by the Borel isomorphism theorem, there exists* $\xi_A : [0, \frac{1}{2}) \leftrightarrow A$*, that is one-to-one, onto, measurable, with a measurable inverse so that* $\{p_r : r \in [0, \frac{1}{2})\} \subset \Pi^\dagger$ *is descriptively complete. Pick* $\mu \in A$ *and a function* $g : [0,1] \to \mathbb{X}$ *such that* $g(\lambda) = \mu$*. If* $f(\frac{1}{2}, u) = f(1, u) = g(u)$*, then* $f(\{\alpha p_{\frac{1}{2}} + (1-\alpha)p_1 : \alpha \in [0,1]\} = \{\mu\}$*, showing that* $\Pi^\dagger$ *is descriptively complete.*

Definition 5. *A measurable* $\Pi \subset \Delta_{\mathcal{F}}$ *is* ***broad sense measurably identifiable*** *if there exists an* $E \in \mathcal{F}$*, and a measurable, onto correspondence* $\Phi : E \rightrightarrows \Pi$ *with* $\{\Phi(\omega) : \omega \in E\}$ *an uncountable measurable partition of* Π *and for all* $p \in \Pi$*,* $p(\Phi^{-1}(p)) = 1$*.*

If Π is broad sense measurably identifiable, then for all $\omega, \omega' \in E$, $\Phi(\omega) = \Phi(\omega')$ or $\Phi(\omega) \cap \Phi(\omega') = \emptyset$, and for $p, p' \in \Pi$, we write $p \sim_\Phi p'$ if $p(\Phi^{-1}(p') = p'(\Phi^{-1}(p) = 1$.

Corollary 2.4. *If* S *is a descriptively complete set of probabilities then there exists an* $E \in \mathcal{F}$ *and a measurable onto correspondence* $\Phi : E \rightrightarrows S$ *for which* $\{\Phi(\omega) : \omega \in E\}$ *is an uncountable measurable partition of* Π*, and for every* $A \in \mathbb{K}_{\Delta(\mathbb{X})}$ *there exist a measurable* $f_A : \Omega \to \mathbb{X}$ *such that* $f_A(S) = A$ *and for all* $p, p' \in S$*,* $[p \sim_\Phi p'] \Rightarrow [f_A(p) = f_A(p')]$*.*

Proof. Because S is descriptively complete, there exists a measurable $f_\circ : \Omega \to [0,1] \times [0,1]$ such that $f_\circ(S)$ is equal to the set Π° from Example 8. For each $r \in [0,1]$, let $S_r = \{p \in S : p(f_\circ^{-1}(\{r\} \times [0,1])) = 1\}$, define $h(\omega) = \text{proj}_1(f_\circ(\omega))$ as the projection of $f_\circ(\omega) \in [0,1]^2$ onto its first component, and define $\Phi(\omega) = S_{h(\omega)}$. Pick $A \in \mathbb{K}_{\Delta(\mathbb{X})}$. Because Π° is descriptively complete, there exists an $h_A : [0,1]^2 \to \mathbb{X}$ such that $h_A(\Pi^\circ) = A$. Define $f_A(\omega) = h_A(f_\circ(\omega))$. □

4.3. **Combining Descriptive Completeness and State Independence.** Suppose that $f, g : \Omega \to \mathbb{X}$ and define the corresponding $f', g' : S \to \mathbb{R}$ by $f'(p) = \int_\Omega \boldsymbol{u}(f(\omega))\, dp(\omega)$ and $g'(p) = \int_\Omega \boldsymbol{u}(g(\omega))\, dp(\omega)$ where $\boldsymbol{u}$ is the decision maker's vNM utility function for risky problems. From GMM $f', g' : S \to \mathbb{R}$ being positive affine transformations of each other implies that $\alpha_{f'} = \alpha_{g'}$.

Definition 6. *We say that multiple prior preferences,* $\succsim$*, over measurable functions* $f : \Omega \to \mathbb{X}$ *are* ***state independent*** *or* ***neutral*** *if* $[f(S) = f'(S)] \Rightarrow [f \sim f']$*.*

In particular, if $\varphi : \Omega \leftrightarrow \Omega$ has the property that $\varphi(S) = S$, then $f \sim f \circ \varphi$.

An implication of the following is that if the preferences over random variables satisfy state independence and GMM's results for α-MEU preferences extend to functions taking on more than finitely many values, then the α must be constant.

Proposition 1. *If* $f, g : \Omega \to \mathbb{X}$*,* $v \in C_b(\mathbb{X})$*,* S *is a descriptively complete set of priors on a standard space, and the mappings* $p \mapsto \int_\Omega v(f(\omega))\, dp(\omega)$ *and* $p \mapsto \int_\Omega v(g(\omega))\, dp(\omega)$ *from* S *to* $\mathbb{R}$ *are not constant, then there exists* $f', g' : \Omega \to \mathbb{X}$ *such that* $f'(S) = f(S)$*,* $g'(S) = g(S)$*, and the mappings* $p \mapsto \int_\Omega v(f'(\omega))\, dp(\omega)$ *and* $p \mapsto \int_\Omega v(g'(\omega))\, dp(\omega)$ *from* S *to* $\mathbb{R}$ *are positive affine transformations of each other.*

Proof. Let $A = f(S)$ and $B = g(S)$. For $x, y \in [0,1]$ define $p_{x,y}$ to be the uniform distribution on $\{(x,y)\} \times [0,1]$. The set $\{p_{x,y} | (x,y) \in [0,1] \times [0,1]\}$ is compact. Let $\tilde{P}$ denote the closure of its convex hull, that is $\tilde{P} = \overline{\text{co}}\{p_{x,y} | (x,y) \in [0,1] \times [0,1]\}$. From [19, Theorem 1 and §3.2], there exists a measurable $R : \Omega \to [0,1]^3$ such that $\{R(p) : p \in C\} = \tilde{P}$.

Since A and B are compact convex sets, for the vNM utility function $v : M \to [0,1]$, the ranges of the maps $p \mapsto \langle v, f(p) \rangle$ and $p \mapsto \langle v, g(p) \rangle$ without loss in generality can be taken as non-degenerate intervals, $[a,b], [a',b'] \subset [0,1]$.

For $r \in [a,b]$, define a closed convex subset of A by $A_r = \{\nu \in A : \langle v, \nu \rangle = r\}$. The correspondence

$$H_A(r, \mu) = \begin{cases} \{(r, \mu)\} & \text{if } \mu \in A_r, \\ \{(r, \nu) | \nu \in A_r\} & \text{else.} \end{cases} \tag{11}$$

from $[a,b]\times\Delta(\mathbb{X})$ to $[a,b]\times\Delta(\mathbb{X})$ is closed valued, convex valued, lower hemicontinuous, defined on a metric, hence paracompact, space, and takes values in a topologically complete, locally convex, vector space. By [40, Theorem 3.2″ (p. 364 *et seq.*)] H_A has a continuous selection h_A with the property that $h_A(r,\mu)\in A_r$.

Let $g:[0,1]\leftrightarrow\Delta(\mathbb{X})$ be a measurable isomorphism and define $f_A(x,g(y))=(r_x,h_A(r_x,g(y)))$, where $r_x:=a+(b-a)x$. The function $f_A(\cdot,\cdot)$ is jointly measurable. This follows from: the measurability of r_x since it is continuous and that of $g(y)$ by its construction; and joint continuity of h_A, which preserves measurability. Moreover, by the construction of the Borel isomorphism $g(\cdot)$ and that of h_A, for each $x\in[0,1]$, $f_A(x,\cdot):[0,1]\to A_{r_x}$ is onto. Therefore $A=\{A_{r_x}|x\in[0,1]\}$. Define the measurable function $\tilde{X}(x,y,z)=\mathbf{b}(h_A(r_x,g(y)),z)$ where $\mathbf{b}(\cdot)$ is the Blackwell-Dubins function. Define the function $f'(\omega)=\tilde{X}(R(\omega))$ as a composite of measurable functions. Observe that $f'(S)=\tilde{X}(R(S))=A$.

Since $[a,b]$ and $[a',b']$ are non-degenerate intervals, $[a',b']=\alpha[a,b]+\beta$ with $\alpha=\frac{b'-a'}{b-a}>0$ and $\beta=\frac{a'b-b'a}{b-a}$. Therefore, $[a',b']=\{s_x=\alpha r_x+\beta,x\in[0,1]\}$. By analogous arguments made in the previous two paragraphs B in the role of A and s_x in the role of r_x we construct $\tilde{Y}(x,y,z)=\mathbf{b}(h_B(s_x,g(y)),z)$ and $g'(\omega)=\tilde{Y}(R(\omega))$ such that $g'(S)=\tilde{Y}(R(S))=B$. By construction, for each $p\in S$, $\langle v,g'(p)\rangle=\alpha\langle v,f'(p)\rangle+\beta$. □

Much less than descriptive completeness may force α_f to be constant in the presence of state independence.

Example 10. *Suppose that $S=\{q_\beta:\beta\in[0,1]\}$ where $q_\beta=\beta p+(1-\beta)p'$, $p\neq p'$ are non-atomic probabilities on $(\Omega,\mathcal{F})$, and $p'=\varphi(p)$ for some $\varphi:\Omega\leftrightarrow\Omega$. Suppose also that $\mathbb{X}=\{a,b,c\}$ and $u:\mathbb{X}\to[0,1]$ satisfies $0=u(a)<u(b)<u(c)=1$. Every $f(S)$ is either 0- or 1-dimensional, so that the mapping $\beta\mapsto\langle u,f(q_\beta)\rangle$ is affine, either constant, increasing, or decreasing. Further, setting $f'=f\circ\varphi$ delivers $f(S)=f'(S)$ but the mapping $\beta\mapsto\langle u,f'(q_\beta)\rangle$ has -1 times the slope. Combined with state independence, the mapping $f\mapsto\alpha_f$ must be constant.*

A variation on the arguments for Proposition 1 yield the following.

Proposition 2. *If $f,g:\Omega\to\mathbb{X}$, $u\in C_b(\mathbb{X})$, S is a descriptively complete set of priors on a standard space, and the sets $\{\int_\Omega u(f(\omega))\,dp(\omega):p\in S\}$ are $\{\int_\Omega u(g(\omega))\,dp(\omega):p\in S\}$ equal, then there exists $f',g':\Omega\to\mathbb{X}$ such that $f'(S)=f(S)$, $g'(S)=g(S)$, and the mappings $p\mapsto\int_\Omega u(f'(\omega))\,dp(\omega)$ and $p\mapsto\int_\Omega u(g'(\omega))\,dp(\omega)$ from S to $\mathbb{R}$ are equal to each other.*

This result has a strong implication for the MBA preferences of [14]. Suppose that preferences over measurable functions from Ω to $\mathbb{X}$ taking on more than finitely many values can be represented by

$$MBA(f)=\alpha_f\cdot\min_{p\in S}\int_\Omega \boldsymbol{u}(f(\omega))\,dp(\omega)+(1-\alpha_f)\cdot\max_{q\in S}\int_\Omega \boldsymbol{u}(f(\omega))\,dq(\omega) \tag{12}$$

where $\alpha_f=\alpha_g$ if the mappings $p\mapsto\int_\Omega \boldsymbol{u}(f'(\omega))\,dp(\omega)$ and $p\mapsto\int_\Omega \boldsymbol{u}(g'(\omega))\,dp(\omega)$ from S to $\mathbb{R}$ are equal to each other. Then, if the preferences are state independent and the set of priors is descriptively complete, then Proposition 2 implies that $\alpha_f=\alpha_g$ if the sets $\{\int_\Omega \boldsymbol{u}(f(\omega))\,dp(\omega):p\in S\}$ are $\{\int_\Omega \boldsymbol{u}(g(\omega))\,dp(\omega):p\in S\}$ equal,

4.4. **Descriptive Incompleteness.** Let S be a set of priors on $(\Omega,\mathcal{F})$ and define $\mathcal{R}(S)=\{A\in\mathbb{K}_{\Delta(\mathbb{X})}:A=f(S),\ f:\Omega\to\mathbb{X}\text{ measurable}\}$. For e.g. $\mathbb{X}=[0,M]$, we have the following.

Theorem 3. *If every non-empty neighborhood in $\mathbb{X}$ is uncountable and S is a set of priors that fails to be descriptively complete, then there is a dense subset of $\mathbb{K}_{\Delta(\mathbb{X})}$ that does not belong to $\mathcal{R}(S)$, equivalently, if $\mathcal{R}(S)$ contains any non-empty, open subset of $\mathbb{K}_{\Delta(\mathbb{X})}$, then S must be descriptively complete.*

Proof. If a descriptively complete subset of $\mathbb{K}_{\Delta(\mathbb{X})}$ belongs to $\mathcal{R}(S)$, then S is descriptively complete (by composition of measurable functions). The proof will be complete once we show that descriptively complete subsets of $\mathbb{K}_{\Delta(\mathbb{X})}$ are dense. Pick arbitrary $A \in \mathbb{K}_{\Delta(\mathbb{X})}$ and $\epsilon > 0$, and let A_f denote a finite set of extreme points for A such that $d(A, \mathbf{co}\,(A_f)) < \epsilon/2$.

For any $\delta > 0$ and $x \in \mathbb{X}$, let $B_\delta(x)$ denote the necessarily uncountable, open ball with radius $\delta > 0$ around $x \in \mathbb{X}$. By the Borel isomorphism theorem, there exists $\varphi_{x,\delta} : [0,1] \times [0,1] \leftrightarrow B_\delta(x)$ where $\varphi_{x,\delta}$ is a measurable bijection with measurable inverse. Let $\Pi_{x,\delta}$ denote the descriptively complete set $\varphi_{x,\delta}(\Pi^\circ)$ where Π° is the descriptively complete set of probabilities from Example 8.

Pick $\delta < \epsilon/2$ such that the points in A_f are at least 2δ from each other. Since the support sets are disjoint, the closed convex hull of the set $\cup_{x \in A_f} \Pi_{x,\delta}$ is descriptively complete and within ϵ of A. □

There are many situations in which the conclusion of Theorem 3 understates the degree to which failing to be descriptively complete limits the set of problems that the decision maker can conceive of.

a. If $S_{c,d} = \{p \in \Delta([0,1]) : c \leq dp/d\lambda \leq d\}$, $0 < c < 1 < d$, then $\mathcal{R}(S_{c,d})$ is a connected union of three line segments if $\#\mathbb{X} = 2$ as in Figure 1(b), while $\mathcal{R}(S_{0,\infty})$ contains only three sets. More generally, for any non-atomic Q, if $S_{c,d}(Q) := \{p \in \Delta(\Omega) : c \leq dp/dQ \leq d\}$, then $\mathcal{R}(S_{c,d}(Q))$ is a negligible subset of $\mathbb{K}_{\Delta(\mathbb{X})}$ if $\mathbb{X}$ is finite, and is a closed 1-shy subset of $\mathbb{K}_{\Delta(\mathbb{X})}$ if $\mathbb{X}$ is infinite.[7]
b. Suppose that ν is a strictly convex distortion of a probability, that is, $\nu(E) := \varphi(P(E))$, $\varphi : [0,1] \to [0,1]$ a strictly convex, increasing, and onto function. If $S = \{p \in \Delta_{\mathcal{F}} : (\forall E)[p(E) \geq \nu(E)]\}$ is the core of ν, then in the two consequence case, $\mathcal{R}(S)$ is the set of intervals of the form $[\varphi(r), 1 - \varphi(1-r)]$, $r \in [0,1]$. Only a one dimensional curve in the two-dimensional set of intervals can be modeled, and for any non-trivial consequence space, $\mathcal{R}(S)$ contains no singleton sets $A = \{\mu\}$ unless μ is a point mass.
c. If $S = \Delta(\Omega)$ and $\mathbb{X}$ is finite, then $\mathcal{R}(S) = \{\Delta(F) : \emptyset \neq F \subset \mathbb{X}\}$, that is, $\mathcal{R}(S)$ contains only the faces of $\Delta(\mathbb{X})$.

5. Separations of Risk and Ambiguity Attitudes

A continuous linear functional, L, on $\mathbb{K}_{\Delta(\mathbb{X})}$, restricted to the closed, convex class of singleton sets, gives an expected utility function, the resultant. By definition, this utility function contains a decision maker's attitude toward risk. The value of L on the rest of $\mathbb{K}_{\Delta(\mathbb{X})}$ contains the attitude toward ambiguity. The only remaining issue is the representation of this part of L. The essential device is a continuous direct sum decomposition of elements of $\mathbb{K}_{\Delta(\mathbb{X})}$ into singleton sets plus sets centered at 0. We do this for two different classes of compact convex sets, $\mathbb{K}^{sym}_{\Delta(\mathbb{X})}$, the centrally symmetric sets, and $\mathbb{K}^{fin}_{\Delta(\mathbb{X})}$, the finite dimensional sets. $\mathbb{K}^{sym}_{\Delta(\mathbb{X})}$ is convex, closed, hence complete, and is nowhere dense in $\mathbb{K}_{\Delta(\mathbb{X})}$ unless $\#\mathbb{X} = 2$. By contrast, $\mathbb{K}^{fin}_{\Delta(\mathbb{X})}$ is convex, equal to $\mathbb{K}_{\Delta(\mathbb{X})}$ if $\#\mathbb{X}$ is finite, and dense in $\mathbb{K}_{\Delta(\mathbb{X})}$ when $\#\mathbb{X}$ is infinite.

A vector space $\mathfrak{X}$ can be expressed as a continuous direct sum, written $\mathfrak{X} = \mathfrak{X}_1 \oplus \mathfrak{X}_2$, if $\mathfrak{X}_1$ and $\mathfrak{X}_2$ are vector subspaces of $\mathfrak{X}$, every $\mathbf{x} \in \mathfrak{X}$ has a unique expression as $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ where $\mathbf{x}_1 \in \mathfrak{X}_1$, $\mathbf{x}_2 \in \mathfrak{X}_2$, and the mappings $\mathbf{x} \mapsto \mathbf{x}_i$ are continuous and linear. Further, given a continuous direct sum, any continuous linear functional on $\mathfrak{X}$ can be expressed as $L(\mathbf{x}) = L(\mathbf{x}_1) + L(\mathbf{x}_2)$ where each $\mathbf{x}_i \mapsto L(\mathbf{x}_i)$ is a continuous linear functional on $\mathfrak{X}_i$, and any pair of continuous linear functionals, L_1 and L_2 on $\mathfrak{X}_1$ and $\mathfrak{X}_2$ can be combined by defining $L(\mathbf{x}) = L_1(\mathbf{x}_1) + L_2(\mathbf{x}_2)$.

We are interested in direct sum decompositions when $\mathfrak{X}$ contains $\mathbb{K}^{fin}_{\Delta(\mathbb{X})}$ and/or $\mathbb{K}^{sym}_{\Delta(\mathbb{X})}$. When the space $\mathbb{X}$ is finite, $\mathbb{K}_{\Delta(\mathbb{X})} = \mathbb{K}^{fin}_{\Delta(\mathbb{X})}$, and we present a complete theory. When the space $\mathbb{X}$ is infinite, decompositions of $\mathbb{K}_{\Delta(\mathbb{X})}$ are more difficult.

[7] A set $E \subset \mathbb{K}_{\Delta(\mathbb{X})}$ is 1-shy if there exists a non-degenerate line segment, $L \subset \mathbb{K}_{\Delta(\mathbb{X})}$, with every translate of E intersection L in a Lebesgue null set. See [4] for a full development of shy subsets of infinite dimensional convex sets.

5.1. **Centrally Symmetric Elements of $\mathbb{K}_{\Delta(\mathbb{X})}$.** Line segments, parallelotopes, and ellipses are classic examples of centrally symmetric sets. Non-degenerate triangles are not centrally symmetric.

Definition 7. *An $A \in \mathbb{K}_{\Delta(\mathbb{X})}$ is **centrally symmetric** if there exists a **symmetric center**,* $\mathrm{Symm}(A) \in A$ *and* $(A - \mathrm{Symm}(A)) = -(A - \mathrm{Symm}(A))$. *The class of centrally symmetric elements in $\mathbb{K}_{\Delta(\mathbb{X})}$ is denoted $\mathbb{K}^{sym}_{\Delta(\mathbb{X})}$.*

The symmetric center mapping, $A \mapsto \mathrm{Symm}(A)$, is linear, Lipschitz continuous. Further, the class $\mathbb{K}^{sym}_{\Delta(\mathbb{X})}$ is a closed, convex subset of $\mathbb{K}_{\Delta(\mathbb{X})}$ that is nowhere dense if $\#\mathbb{X} \geq 3$.

5.2. **Finite Dimensional Elements of $\mathbb{K}_{\Delta(\mathbb{X})}$.** The most extensively used and studied continuous linear center for compact and convex sets that are not centrally symmetric is the Steiner point. It agrees with the symmetric center when the latter exists, but, unlike the symmetric center, does not have a continuous extension to infinite dimensional sets. The setting for Steiner points is the class of finite dimensional subsets of a separable Hilbert space $\mathbb{H}$. We fix a continuous linear embedding of $\Delta(\mathbb{X})$ into $\mathbb{H}$, and let $\mathbb{K}^{fin}_{\Delta(\mathbb{X})}$ denote the class of finite dimensional compact and convex subsets of $\mathbb{H}$.[8]

Definition 8. *The **Steiner point** of a finite dimensional, compact, convex $A \subset \mathbb{H}$ is the vector-valued integral*

$$\mathrm{St}(A) = \int_{S^{\ell-1}} \operatorname*{argmin}_{\mu \in A} \langle h, \mu \rangle \, d\lambda_\ell(h) \tag{13}$$

where V_ℓ is an ℓ-dimensional subset of $\mathbb{H}$ containing A and λ_ℓ is the uniform distribution on $\{h \in V_\ell : \|h\| = 1\}$.

Since A is convex, the mass of λ_ℓ is 1, and λ_ℓ has full support on $\{h \in V_\ell : \|h\| = 1\}$, $\mathrm{St}(A)$ is in the relative interior of A. Linearity of $A \mapsto \mathrm{St}(A)$ is immediate. The theorem of the maximum tells us that for each f, $A \mapsto \operatorname{argmin}_{\mu \in A} \langle f, \mu \rangle$ is upper hemi-continuous. For all but a λ-null subset of f, the argmin is single-valued, so that dominated convergence implies the continuity of $A \mapsto \mathrm{St}(A)$. Among the continuous linear centers, the Steiner point is uniquely determined by the property that it commutes with rigid motions (see e.g. [49, Theorem 3.4.2]).

For finite dimensional centrally symmetric sets, the center of symmetry and the Steiner point agree, and the Steiner point extends continuously to the infinite dimensional elements of $\mathbb{K}^{sym}_{\Delta(\mathbb{X})}$. Restricted to ℓ-dimensional sets, the Steiner point mapping is Lipschitz continuous with a Lipschitz constant that goes to ∞ as $\ell \uparrow \infty$. Thus, if a sequence, $A_n \to A$ and the A_n are of bounded dimensionality, the dimensionality of A has the same bound, and $\mathrm{St}(A_n) \to \mathrm{St}(A)$.

The Steiner point mapping cannot be continuously extended from the dense $\mathbb{K}^{fin}_{\Delta(\mathbb{X})}$ to $\mathbb{K}_{\Delta(\mathbb{X})}$ in $\mathbb{H}$. Vitale [55] shows that for any infinite dimensional Hilbert space, every compact convex A has the following property: if $x \in A$, then $x = \lim_n \mathrm{St}(A_n)$ where A_n is a sequence of finite dimensional subsets of $\mathbb{H}$ with dimensionality going to ∞ and $d_H(A, A_n) \to 0$. This means that there is no way to give a continuous linear decomposition of the compact convex subsets of $\mathbb{H}$ that agrees with the Steiner point decomposition restricted to finite dimensional sets.

Vitale's argument depends on the existence of highly asymmetric finite dimensional sets pointing in 'every' direction. This is automatically satisfied in $\mathbb{H}$, but, in principle, if $\mathbb{X}$ is compact and A and the A_n are constrained to be subsets of the compact convex $\Delta(\mathbb{X}) \subset \mathbb{H}$, then the argument might not go through. But it does, at least partly.[9]

[8] With $\mathbb{X} = [0, M]$, Machina [37] identified every $\mu \in \Delta(\mathbb{X})$ with its cumulative distribution function, F_μ, and used the L^2 distance $\|F_\mu - \mathcal{F}_\nu\|_2 = \left(\int_0^M |F_\mu(x) - \mathcal{F}_\nu(x)|^2 \, dx\right)^{1/2}$. For more general metric spaces, similar embeddings are available. If $\mathbb{X}$ is finite, we can take $\mathbb{H}$ to be $\mathbb{R}^{\mathbb{X}}$ with the usual 2-norm.

[9] Sketchily, the proof of the following takes the e_n to have μ's density plus terms of the form $\epsilon(1_{[k/n,(k+1)/n)} - 1_{[k'/n,(k'+1)/n)})$, $k \neq k'$, in Vitale's proof.

Lemma 1. *With* $\mathbb{X} = [0, M]$*, under the embedding of* $\mu \in \Delta([0, M])$ *as its cdf,* $\mu \mapsto F_\mu \in L^2[0, M]$*, for every* $A \in \mathbb{K}_{\Delta(\mathbb{X})}$*, if* $\mu \in \mathrm{extr}(A)$ *is mutually absolutely continuous with respect to Lebesgue measure, then there exists a sequence of finite dimensional* A_n *in* $\mathbb{K}_{\Delta(\mathbb{X})}$ *such that* $d_H(A_n, A) \to 0$ *and* $\mathrm{St}(A_n) \to \mu$*.*

5.3. **Decompositions and Ambiguity Aversion.** $\mathbb{K}^{sym}_{\Delta(\mathbb{X})}$ has the direct sum decomposition into $\mathbb{H} \oplus \mathrm{Symm}^{-1}(0)$ and $\mathbb{K}^{fin}_{\Delta(\mathbb{X})}$ has the direct sum decomposition into $\mathbb{H} \oplus \mathrm{St}^{-1}(0)$,

$$(\forall A \in \mathbb{K}^{sym}_{\Delta(\mathbb{X})})[A = \mathrm{Symm}(A) + (A - \mathrm{Symm}(A))], \text{ and} \tag{14}$$

$$(\forall A \in \mathbb{K}^{fin}_{\Delta(\mathbb{X})})[A = \mathrm{St}(A) + (A - \mathrm{St}(A))]. \tag{15}$$

According to context, let $\mathrm{Cent}(A)$ denote $\mathrm{Symm}(A)$ if A is centrally symmetric, and let it denote $\mathrm{St}(A)$ if A is finite dimensional. In terms of the representation in Theorem 1, the resultant $\boldsymbol{u}$ of L captures the attitude toward risk, while the restriction of L to the vector subspace $\mathrm{Cent}^{-1}(0)$ captures attitudes toward ambiguity. The following records the separation result.

Proposition 3. *The restriction of every continuous linear* $L : \mathbb{K}_{\Delta(\mathbb{X})} \to \mathbb{R}$ *to* $\mathbb{K}^{sym}_{\Delta(\mathbb{X})}$ *or to* $\mathbb{K}^{fin}_{\Delta(\mathbb{X})}$ *is, up to positive affine transformation, of the form*

$$L(A) = \langle \boldsymbol{u}, \mathrm{Cent}(A) \rangle + L(A - \mathrm{Cent}(A)) \tag{16}$$

where $\boldsymbol{u}$ *is the resultant of* L*.*

Let L° denote the restriction of L to the vector subspace, $\mathrm{Cent}^{-1}(0)$, so that $(\boldsymbol{u}, L^\circ)$ is the direct sum decomposition of a continuous linear L (on either $\mathbb{K}^{sym}_{\Delta(\mathbb{X})}$ or $\mathbb{K}^{fin}_{\Delta(\mathbb{X})}$). Combining an arbitrary $\boldsymbol{u}$ with an arbitrary L° delivers a continuous linear functional. Ambiguity aversion is dislike of expansions of a set around its center. In the two outcome case discussed in §2, this is the negativity of $L^\circ([-r, +r]) = -vr$ for $r \geq 0$. The following is the direct generalization.

Definition 9. *For* $\mathcal{A} \subset \mathbb{K}^{sym}_{\Delta(\mathbb{X})}$ *or* $\mathcal{A} \subset \mathbb{K}^{fin}_{\Delta(\mathbb{X})}$*, a continuous linear* $(\boldsymbol{u}, L^\circ)$ *is*

(1) ***ambiguity averse relative to*** $\mathcal{A}$ *if* $L^\circ(\mathrm{Cent}^{-1}(0) \cap \mathcal{A}) \subset \mathbb{R}_-$*,*
(2) ***ambiguity neutral relative to*** $\mathcal{A}$ *if* $L^\circ(\mathrm{Cent}^{-1}(0) \cap \mathcal{A}) = \{0\}$*, and*
(3) ***ambiguity loving relative to*** $\mathcal{A}$ *if* $L^\circ(\mathrm{Cent}^{-1}(0) \cap \mathcal{A}) \subset \mathbb{R}_+$*.*

Preferences $(\boldsymbol{u}_1, L_1^\circ)$ *are* ***more ambiguity averse relative to*** $\mathcal{A}$ ***than those given by*** $(\boldsymbol{u}_2, L_2^\circ)$ *if* $L_1^\circ(A - \mathrm{Cent}(A)) \leq L_2^\circ(A - \mathrm{Cent}(A))$ *for all* $A \in \mathcal{A}$*.*

Comments.

a. The comparison between degrees of ambiguity aversion does not depend on $\boldsymbol{u}_1$ or $\boldsymbol{u}_2$. If $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ belong to $\mathbb{U}_0^1$ in Definition 9, as they would if they were comonotonic, then they have the same sup and inf so that the comparison of L_1° and L_2° is on the same scale as the vNM utility.
b. If $\#\mathbb{X} \geq 3$, then for any ambiguity averse L_2°, the set of L_1° that are more ambiguity averse is an infinite dimensional cone. Olszewski [42] gives a definition of ambiguity aversion for preferences over $\mathbb{K}_{\Delta(\mathbb{X})}$ in which the set $\mathcal{A}$ depends on $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, and the set of more ambiguity averse preferences is one dimensional.
c. Complete separation of ambiguity and risk attitudes can be achieved by other means: Gajdos et al. [24] give a different description of choice problems, hence a different domain for preferences, one that allows for complete separation; Klibanoff et al. [33] give preferences over random variables that are not determined by a set of induced distributions (see §6.5).

6. Old Puzzles and New Preferences

Continuous linear preferences on $\mathbb{K}_{\Delta(\mathbb{X})}$ illuminate a number of old puzzles and contain new classes of preferences.

Puzzles

- Comparative ambiguity aversion. Extant theories of comparative ambiguity aversion for preferences over random variables has been, essentially, limited to comparing the ambiguity attitudes of decision makers with the same attitude toward risk. Part of the difficulty with making comparisons across people has been a neglect of the role that a set of priors plays in determining the set of problems that a decision maker can conceive of — if two sets of priors both fail descriptive completeness, it may be that the only decision problems that the two decision makers can commonly contemplate are trivial.[10] Proposition 5 shows that this is only part of the problem: α-MEU preferences are ambiguity averse with respect to the class of centrally symmetric sets iff $\alpha \geq \frac{1}{2}$ and ambiguity aversion relative to this set is increasing in α; however, α-MEU preferences fail to be ambiguity averse relative to any class $\mathcal{A}$ that includes the triangular sets of distributions unless $\alpha = 1$.
- Constant relative ambiguity aversion. One can study the effects of wealth on risk behavior, e.g. portfolio choices, by studying how ratios of the first and second derivative of the expected utility function behave at higher levels of wealth. This means that within the class of linear-in-probabilities expected utility preferences, one can observe increasing or decreasing absolute or relative risk aversion. On simple classes of sets, the RDU, CEU, and α-MEU preferences demonstrate constant ambiguity aversion, and Machina [38] gives several counter-intuitive examples that arise from this constancy. We show that constancy is not a necessary feature of preferences that are linear in sets, and show how this allows us to resolve Machina's challenges. In a simple model of loss insurance, we show that neglect of decreasing relative ambiguity aversion will bias measures of decreasing risk aversion.
- Updating sets of probabilities. As an illustration of the use of being able to update in general information structures, we present a solution to the problem of convincing an ambiguous Bayesian.

New classes of preferences

- Decreasing/increasing relative ambiguity aversion. We specify linear-in-sets preferences for a subset, $\mathcal{Z}_{\mathbb{X}} \subset \mathbb{K}^{sym}_{\Delta(\mathbb{X})}$, with decreasing (or increasing) relative ambiguity aversion. A simple insurance demand problem demonstrates that, in the presence of ambiguity, measured decreases of relative risk aversion may be upwards biased.
- Utility range dominance. Corollaries 1.1 and 1.2 showed that if $\eta_{\min}$ and $\eta_{\max}$ are supported on the set of non-decreasing (resp. non-decreasing and concave functions), then preferences respect first (resp. second) order stochastic dominance relations. Support sets give rise to a novel form of dominance, *utility range dominance*, that can distinguish between sources of ambiguity.

NB: for ease, in this section we assume that $\mathbb{X} = [0, M]$.

6.1. **Orderings of $\eta_{\min}$ and $\eta_{\max}$.** Positive measures are ordered by $\eta_{\min} \geq \eta_{\max}$ if for all measurable $E \subset \mathbb{U}^1_0$, $\eta_{\min}(E) \geq \eta_{\max}(E)$. We first examine what can be said in the general case, then turn to the case where $\eta_{\min}$ and $\eta_{\max}$ put all their mass on singletons.

6.1.1. *The General Case.* A sufficient condition for ambiguity aversion relative to $\mathbb{K}^{sym}_{\Delta(\mathbb{X})}$ is that the "pessimistic part" of the preferences, $\eta_{\min}$, is larger than the "optimistic part," $\eta_{\max}$. A sufficient condition for ambiguity aversion relative to $\mathbb{K}^{fin}_{\Delta(\mathbb{X})}$ is that there be no "optimistic part." For the special case of α-MEU preferences, we will see that these condition are also necessary.

Proposition 4. *If $\eta'_{\min} \geq \eta_{\min} \geq \eta_{\max} \geq \eta'_{\max}$, then L_η is ambiguity averse relative to $\mathbb{K}^{sym}_{\Delta(\mathbb{X})}$ and $L_{\eta'}$ is more ambiguity averse. If $\eta_{\max} = 0$, then L_η is ambiguity averse relative to $\mathbb{K}^{fin}_{\Delta(\mathbb{X})}$.*

[10] We conjecture that this is a generic property of pairs of sets of priors failing descriptive completeness.

Proof. Suppose first that $\eta'_{\min}(E) \geq \eta_{\min}(E) \geq \eta_{\max}(E) \geq \eta'_{\max}(E)$ for any measurable E. For any $u \in \mathbb{U}^1_0$ and $A_0 \in \text{Symm}^{-1}(0)$, $\psi_{A_0}(u) \leq 0 \leq -\psi_{A_0}(u) = \psi^{A_0}(u)$. Therefore,

$$\int \psi_{A_0}(u)\, d\eta'_{\min}(u) + \int \psi^{A_0}(u)\, d\eta'_{\max}(u) \leq \\ \int \psi_{A_0}(u)\, d\eta_{\min}(u) + \int \psi^{A_0}(u)\, d\eta_{\max}(u) \leq 0. \tag{17}$$

Because $A_0 \in \text{Symm}^{-1}(0)$, both sides of the inequality in (17) are linear in $r \geq 0$ when A_0 is replaced with $r \cdot A_0$, proving the first assertion.

Now suppose that $\eta_{\max} = 0$. For the second part of the Proposition, note that for any A containing 0 in its algebraic interior and any $u \in \mathbb{U}^1_0$, $\psi_A(u) \leq 0$. Any $A_0 \in \text{St}^{-1}(0)$ contains 0 in its algebraic interior. Therefore, $L^\circ(r \cdot A_0) = \int \psi_{r \cdot A_0}(u)\, d\eta_{\min}(u)$ is a decreasing function of $r \geq 0$. □

6.1.2. *α-MEU and Related Preferences.* Theorem 3 tells us that if a decision maker is modeled as having α-MEU preferences but does not have a descriptively complete set of priors, then the model precludes the decision maker from considering a large subset of the problems that may be of interest. From Proposition 1, we know that a model of a decision maker having α-MEU preferences and a descriptively complete set of priors must have a constant α. We now examine the question of whether having a constant $\alpha \geq \frac{1}{2}$ corresponds to ambiguity aversion, or whether increases in α correspond to increases in ambiguity aversion. Recall that the line segments joining probability distributions are centrally symmetric, but non-degenerate triangles, i.e. the convex hulls of affinely independent probability distributions are not.

Proposition 5. *Preferences represented by L_η with $\eta = (\eta_{\min}, \eta_{\max}) = (\alpha\delta_{\boldsymbol{v}_1}, (1-\alpha)\delta_{\boldsymbol{v}_2})$ satisfy the following:*

(a) *if $\boldsymbol{v}_1 \neq \boldsymbol{v}_2$, then the preferences are ambiguity averse to $\mathcal{A} \subset \mathbb{K}^{sym}_{\Delta([0,M])}$ containing the line segments iff $\alpha = 1$;*
(b) *if $\boldsymbol{v}_1 = \boldsymbol{v}_2$, then the preferences are ambiguity averse to $\mathcal{A} \subset \mathbb{K}_{\Delta([0,M])}$ containing the triangles iff $\alpha = 1$; and*
(c) *if $\boldsymbol{v}_1 = \boldsymbol{v}_2$, then the preferences are ambiguity averse to $\mathbb{K}^{sym}_{\Delta([0,M])}$ iff $\alpha \geq \frac{1}{2}$, and the ambiguity aversion increases as α does.*

Proof. For any $u \in \mathbb{U}^1_0$, $A \in \mathbb{K}^{sym}_{\Delta([0,M])}$, and $\mu' \in \text{argmin}_{\mu \in A}\langle u, \mu\rangle$, the line through μ' and $\text{Symm}(A)$ goes through a $\mu'' \in \text{argmax}_{\mu \in A}\langle u, \mu\rangle$, and $\text{Symm}(A) = \frac{1}{2}\mu' + \frac{1}{2}\mu''$, completing the first part of the argument.

Now suppose that $\#\mathbb{X} \geq 3$ and that $\mathcal{A} \subset \mathbb{K}^{fin}_{\Delta([0,M])}$ contains the triangles. If $\alpha = 1$, then ambiguity aversion relative to any class of sets is immediate.

Suppose now that $\alpha < 1$. We will show that the preferences given by L_η fail to be ambiguity averse relative to triangles. We require the following.

Lemma 2. *If $T_h = \mathbf{co}\,((-1,0),(0,h),(+1,0))$, $h > 0$, and $(x_h, y_h) = \text{St}(T_h)$, then $\lim_{h \downarrow 0} \frac{h}{y_h} = \infty$.*

Proof. Let $\theta_h = \tan^{-1}(h)$ be the acute angle in T_h, and let U_h be the set of u in the unit circle with $(0,h) \in \text{argmin}_{\mu \in T_h}\langle u, \mu\rangle$. By complementary angles, that $\lambda(U_h) = \frac{\theta_h}{\pi - \theta_h}$. The height of y_h is at most $h \cdot \lambda(U_h) = h \cdot \frac{\theta_h}{\pi - \theta_h}$. Thus, $\frac{h}{y_h} \geq \frac{\pi - \theta_h}{\theta_h} \uparrow \infty$. □

Returning to the proof of Proposition 5, pick ν_1, ν_2, ν_3 such that $\langle u, \nu_1\rangle = \langle u, \nu_2\rangle < \langle u, \nu_3\rangle$, and let Pl denote the intersection of $\Delta(\mathbb{X})$ and the smallest affine space containing ν_1, ν_2, and ν_3. Let lin be an affine map from $\mathbb{R}^2$ to Pl such that $lin((-1,0)) = \frac{3}{4}\nu_1 + \frac{1}{4}\nu_2$, $lin((0,1)) = \frac{1}{4}\nu_1 + \frac{3}{4}\nu_2$, and $\langle u, lin((0,h))\rangle > \langle u, lin((-1,0))\rangle = \langle u, lin((0,1))\rangle$. By Lemma 2, for any $\epsilon > 0$, for sufficiently small h, radial expansions of $A_h := lin(T_h)$ around $\text{St}(lin(T_h))$ decrease $\min_{\mu \in A_h}\langle u, \mu\rangle$ by at most ϵ times the increase in $\max_{\nu \in A_h}\langle u, \nu\rangle$. For any $\alpha < 1$, this implies that for small enough h, radial expansions of A_h increase utility. □

In Proposition 5(c), ambiguity aversion increase as α increases, to put it another way, the set of more ambiguity averse preferences is one-dimensional. With a very different definition of comparative ambiguity aversion, Olszewski [42, Cor. 2] has a similar result.

6.2. **Relative Ambiguity Aversion.** For $\mu, \nu \in \Delta(\mathbb{X})$, we define line segments by $[\![\mu, \nu]\!] := \{(1-\alpha)\mu + \alpha\nu : \alpha \in [0,1]\}$ (note the convention that increases in α put more weight on the right-hand element of the bracketed interval $[\![\mu, \nu]\!]$). For fixed μ, ν, a subset $[a_{\mu,\nu}, b_{\mu,\nu}]$ of $[\![\mu, \nu]\!]$ is defined as $\{(1-\alpha)\mu + \alpha\nu : \alpha \in [a_{\mu,\nu}, b_{\mu,\nu}]\}$, and $[a,b] = [a_{\mu,\nu}, b_{\mu,\nu}]$ is often conveniently rewritten as $[a,b] = [c-r, c+r]$ where the center is $c = \frac{1}{2}(a+b)$ and the radius is $r = \frac{1}{2}(b-a)$. Note that $[a_{\mu,\nu}, b_{\mu,\nu}] = [0,1] = [\![\mu, \nu]\!]$ has center $c = \frac{1}{2}$ and radius $r = \frac{1}{2}$, irrespective of how close together μ and ν are.

6.2.1. *Definition and Examples.* For a continuous linear $L : \mathbb{K}_{\Delta([0,M])} \to \mathbb{R}$ and all $0 \leq x \leq y \leq M$, define $f_L(x,y) = L([\![\delta_x, \delta_y]\!])$ so that $f_L(x,x) = \boldsymbol{u}(x)$, and $f_L(\cdot,\cdot)$ is increasing if L respects first order dominance. By linearity, if $\{\beta_i : i \in I\}$ are a convex set of weights, $L(\sum_i \beta_i([\![\delta_{x_i}, \delta_{y_i}]\!]) = \sum_i \beta_i f_L(x_i, y_i)$. We will investigate the properties of preferences satisfying rank dependence axioms by examining the properties of their associated f_L's.

By the representation theorem and the analysis of the two outcome case in §2.3, $f_L(x,y) = [\frac{1}{2}\boldsymbol{u}(x) + \frac{1}{2}\boldsymbol{u}(y)] - \frac{1}{2}v_{x,y}$ for some $v_{x,y} \in [-1,+1]$ (the $\frac{1}{2}$ arises because the radius of $[\![\delta_x, \delta_y]\!]$ is $\frac{1}{2}$). The number, $v_{x,y}$, gives the degree of ambiguity aversion at the interval $[\![\delta_x, \delta_y]\!]$, and $v_{x,y} \geq 0$ corresponds ambiguity aversion. Scaling $v_{x,y}$ by the utility difference across the interval $[\![\delta_x, \delta_y]\!]$ gives the measure of relative ambiguity aversion.

Definition 10. *For a continuous linear $L : \mathbb{K}_{\Delta([0,M])} \to \mathbb{R}$, **$L$'s relative ambiguity aversion at** $x < y$ is*

$$\rho_{x,y} = \frac{v_{x,y}}{[\boldsymbol{u}(y) - \boldsymbol{u}(x)]}, \tag{18}$$

*and L has **constant/decreasing/increasing relative ambiguity aversion** if $(x,y) \mapsto \rho_{x,y}$ is constant/decreasing/increasing.*

Example 11. *For α-MEU preferences, $f_\alpha(x,y) = \alpha\boldsymbol{u}(x) + (1-\alpha)\boldsymbol{u}(y)$ so that $v_{x,y} = (2\alpha - 1)[\boldsymbol{u}(y) - \boldsymbol{u}(x)]$ and $\rho_{x,y} = (2\alpha - 1)$ is constant.*

Suppose that $x_1, \ldots, x_n \in \mathbb{X}$ have ranked utilities, $\boldsymbol{u}(x_1) \leq \cdots \leq \boldsymbol{u}(x_n)$, that $E_1, \ldots, E_n$ is a measurable partition of Ω, and that $C(\cdot)$ is a capacity on $\mathcal{F}$. The Choquet expect utility of $f := \sum_k x_k 1_{E_k}$ is

$$\begin{aligned} CEU(f) = \boldsymbol{u}(x_1)C(\cup_{k\geq 1}E_k) + [\boldsymbol{u}(x_2) - \boldsymbol{u}(x_1)]C(\cup_{k\geq 2}E_k) \\ + \cdots + [\boldsymbol{u}(x_n) - \boldsymbol{u}(x_{n-1})]C(\cup_{k\geq n}E_k), \end{aligned} \tag{19}$$

that is, succesive increments to utility get lower decision weights. Rewriting yields

$$\begin{aligned} CEU(f) = \boldsymbol{u}(x_1)[C(\cup_{k\geq 1}E_k) - C(\cup_{k\geq 2}E_k)] + \\ \boldsymbol{u}(x_2)[C(\cup_{k\geq 2}E_k) - C(\cup_{k\geq 3}E_k)] + \cdots + \boldsymbol{u}(x_n)[C(\cup_{k\geq n}E_k)]. \end{aligned} \tag{20}$$

Thus, CEU is a monotonic, linear function over ranked vectors of utilities.

Example 12. *For any partition, E_x, E_y of Ω, and any capacity $C(\cdot)$,*

$$L_C([\![\delta_x, \delta_y]\!]) := CEU(x1_{E_x} + y1_{E_y}) = \alpha\boldsymbol{u}(x) + (1-\alpha)\boldsymbol{u}(y) \tag{21}$$

where $\alpha = [C(\Omega) - C(E_y)]$ and $(1-\alpha) = [C(E_y)]$ so that all Choquet expected utility preferences have constant relative ambiguity aversion.

Let $\varphi : [0,1] \to [0,1]$ be increasing, onto, and convex. For a non-atomic P, define the capacity C by $C(E) = \varphi(P(E))$, and let Π be the core of C. The rank dependent expected utility (RDU) of a measurable $f : \Omega \to \mathbb{X}$ is $RDU(f) = \min_{p\in\Pi} \int_\Omega u(f(\omega))\, dp(\omega)$, which is equal to the Choquet integral of f, so that, once again, the two outcome preferences have constant relative ambiguity aversion.

6.2.2. *Decreasing Relative Ambiguity Aversion and Zonoids.* We first characterize the closed, convex subset, $\mathcal{Z}_{\mathbb{X}} \subset \mathbb{K}_{\Delta([0,M])}$, called point mass zonoids, for which specifying a function $(x, y) \mapsto f(x, y)$ determines the preferences. We then turn to functions with $v_{x,y} = (h(x) + h(y))[\boldsymbol{u}(y) - \boldsymbol{u}(x)]$, i.e. functions of the form

$$f(x, y) = [\tfrac{1}{2}\boldsymbol{u}(x) + \tfrac{1}{2}\boldsymbol{u}(y)] - \tfrac{1}{2}(h(x) + h(y))[\boldsymbol{u}(y) - \boldsymbol{u}(x)] \tag{22}$$

where $h : [0, M] \to [0, \frac{1}{2}]$ is a decreasing (or increasing) function. Such functions have relative ambiguity aversion $\rho_{x,y} = (h(x) + h(y))$.

The examples below involve preferences over sets of probabilities of the form $p[\![\delta_{x_1}, \delta_{x_2}]\!] + (1-p)[\![\delta_{x_3}, \delta_{x_4}]\!]$ for different choices of p, x_1, x_2, x_3, and x_4. Such sets belong to the subclass of the centrally symmetric sets known as point mass zonoids.

Definition 11. *$A \in \mathbb{K}^{sym}_{\Delta([0,M])}$ is a **point mass zonotope** if it is a finite convex combination of line segments $[a_{x,y}, b_{x,y}] \subset [\![\delta_x, \delta_y]\!]$ for point masses $\delta_x, \delta_y \in \Delta(\mathbb{X})$. A is a **point mass zonoid** if it belongs to the closure of the set of point mass zonotopes. The class of pointmass zonoids is denoted $\mathcal{Z}_{\mathbb{X}}$.*[11]

A continuous affine functional on $\Delta(\mathbb{X})$ is determined by its values on the extreme points of $\Delta(\mathbb{X})$, that is, on the point masses. Indeed, there is an isometric isomorphism between the continuous affine functionals and $C(\mathbb{X})$, the continuous functions on the set of extreme points. There is an analogous result for the point mass zonoids.

Lemma 3. *If $\mathbb{X}$ is compact, then restriction to the extreme elements of $\mathcal{Z}_{\mathbb{X}}$ defines an isometric isomorphism between the continuous affine functions on $\mathcal{Z}_{\mathbb{X}}$ and the continuous functions $f : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ satisfying $f(x, y) = f(y, x)$.*

Specifically, the extreme points of $\mathcal{Z}_{\mathbb{X}}$ are the 0- and 1-dimensional line segments $[\![\delta_x, \delta_x]\!]$ and $[\![\delta_x, \delta_y]\!]$. If $L : \mathcal{Z}_{\mathbb{X}} \to \mathbb{R}$ is continuous and affine, defining $f_L(x, y) = L([\![\delta_x, \delta_y]\!])$ delivers the continuous function, if f is continuous and symmetric, then defining $L_f([\![\delta_x, \delta_y]\!]) = f(x, y)$ and extending it by continuity and linearity delivers the continuous affine function, the mappings $L \mapsto f_L$ and $f \mapsto L_f$ are bijective inverses of each other, and $\max_{A \in \mathcal{Z}_{\mathbb{X}}} |L(A) - L'(A)| = \max_{x,y \in \mathbb{X}} |f_L(x, y) - f_{L'}(x, y)|$.

Proof. $\mathcal{Z}_{\mathbb{X}}$ is compact and convex, and $\mathrm{extr}(\mathcal{Z}_{\mathbb{X}})$ is closed, hence compact. If τ and τ' are probability distributions on $\mathrm{extr}(\mathcal{Z}_{\mathbb{X}})$ and $\tau \neq \tau'$, one can construct an $L \in \mathcal{L}^{\circ}$ such that $\int L\, d\tau \neq \int L\, d\tau'$ so that $\mathcal{Z}_{\mathbb{X}}$ is a simplex. Bauer's Theorem[12] tells us that if S is a compact simplex, then $\mathrm{extr}(S)$ is closed if and only if the restriction mapping, $a \mapsto a_{|\,\mathrm{extr}(S)}$, defines an isometric isomorphism between $C_{aff}(S)$, the continuous affine functions on S with the sup norm, and $C(\mathrm{extr}(S))$, the continuous functions on $\mathrm{extr}(S)$, also with the sup norm. ☐

6.2.3. *Decreasing Relative Ambiguity Aversion in Loss Insurance.* Suppose that a decision maker has wealth W and faces a loss of size L with probability $(1-p)$ where $p \in [c-r, c+r] \subset [0, 1]$. If L is a continuous linear functional on $\mathcal{Z}_{\mathbb{X}}$, then, letting $\rho = \rho_{W-L,W}$,

$$\begin{aligned} L([a, b]) = [(1-c)\boldsymbol{u}(W-L) + c\boldsymbol{u}(W)] - r\rho[\boldsymbol{u}(W) - \boldsymbol{u}(W-L)] = \\ [(1-(c-r\rho))\boldsymbol{u}(W-L) + (c-r\rho)\boldsymbol{u}(W)] \end{aligned} \tag{23}$$

where $[a, b] = [c-r, c+r]$. The decision maker's certainty equivalent, $X = X(W, L, c, r)$, is defined by $L(\{\delta_X\}) = L([c-r, c+r])$, that is, by

$$X = \boldsymbol{v}((1-(c-r\rho))\boldsymbol{u}(W-L) + (c-r\rho)\boldsymbol{u}(W)) \tag{24}$$

[11] Zonoids in $\mathbb{R}^{\ell}$ are defined as finite convex combinations of line segments, equivalently characterized as (the translate of) the range of non-atomic, $\mathbb{R}^{\ell}$-valued measures [12]. For a comparison of alternative definitions of infinite dimensional polytopes, a class that includes the zonotopic subsets of $\mathbb{K}_{\Delta([0,M])}$, see [44]. All two dimensional centrally symmetric sets are zonoids, but the zonoids are nowhere dense in among centrally symmetric sets of dimension three or higher.

[12] The direct reference is [7, Satz III], or else consult Phelps [45, Proposition 11.1].

where $\boldsymbol{v} := \boldsymbol{u}^{-1}$. Under risk and ambiguity aversion, this means that the certainty equivalent increases in W, and increases more quickly for higher values or r or of ρ. Attributing the increase only to risk aversion corresponds to setting r or ρ to 0, which can lead to overestimating the degree to which sensitivity to risk decreases with W.

6.2.4. *Trading Off Risk and Ambiguity.* Machina [38] gives the following two pairs of ambiguous choice problems with point mass zonotope outcomes based on the four strictly ranked consequences $x_1 < x_2 < x_3 < x_4$, all in $[0, M]$,

$$f_1 = \left(\tfrac{1}{2} - \epsilon\right) \delta_{x_3} + \left(\tfrac{1}{2} + \epsilon\right) \delta_{x_2} \text{ versus} \tag{25}$$

$$f_2 = \left(\tfrac{1}{2} - \epsilon\right) [\![\delta_{x_2}, \delta_{x_3}]\!] + \left(\tfrac{1}{2} + \epsilon\right) [\![\delta_{x_2}, \delta_{x_3}]\!], \text{ and}$$

$$f_3 = \left(\tfrac{1}{2} - \epsilon\right) [\![\delta_{x_3}, \delta_{x_4}]\!] + \left(\tfrac{1}{2} + \epsilon\right) [\![\delta_{x_1}, \delta_{x_2}]\!] \text{ versus} \tag{26}$$

$$f_4 = \left(\tfrac{1}{2} - \epsilon\right) [\![\delta_{x_2}, \delta_{x_4}]\!] + \left(\tfrac{1}{2} + \epsilon\right) [\![\delta_{x_1}, \delta_{x_3}]\!].$$

Option f_2 is equivalent to $[\![\delta_{x_2}, \delta_{x_3}]\!]$, its Steiner point is $\frac{1}{2}\delta_{x_2} + \frac{1}{2}\delta_{x_3}$, while f_1 is the slightly worse distribution just to the "left" of the Steiner point, $(\frac{1}{2}+\epsilon)\delta_{x_2}+(\frac{1}{2}-\epsilon)\delta_{x_3}$. For any positive amount of ambiguity aversion, for all small $\epsilon > 0$, $f_1 \succ f_2$. Machina argues that reasonable preferences might simultaneously have $f_4 \succ f_3$, but that the class of multiple prior preferences satisfying a rank-dependence axiom have $f_1 \succ (\succsim) f_2$ if and only $f_3 \succ (\succsim) f_4$.

Dominance requires that the second part of f_4 dominates the second part of f_3, that is,

$$[\![\delta_{x_1}, \delta_{x_3}]\!] \succ [\![\delta_{x_1}, \delta_{x_2}]\!],$$

while the first part of f_3 dominates the first part of f_4,

$$[\![\delta_{x_3}, \delta_{x_4}]\!] \succ [\![\delta_{x_2}, \delta_{x_4}]\!].$$

Direct calculations show that if relative ambiguity aversion is constant, then $f_1 \succ (\succsim) f_2$ if and only if $f_3 \succ (\succsim) f_4$.

However, if $x_1 \ll x_2 \simeq x_3 \ll x_4$ and decreasing ambiguity aversion is given by a decreasing function $h(\cdot)$ as in (22) with $h(x_1) > h(x_2) \simeq h(x_3) \simeq h(x_4)$, then decreasing relative ambiguity aversion yields both $f_1 \succ f_2$ and $f_3 \prec f_4$ for small ϵ.

6.3. **Updating Sets of Probabilities.** In the following risky problem (based on [32]), we have a cheap talk signaling game [16] with the sender able to commit to a signaling structure.

The Judge/Jury has a prior probability p of the accused person being Guilty and probability $1 - p$ of being Innocent. Judge/Jury can either convict or acquity, and, after normalization their utility function is given by

	Innocent	Guilty
Convict	0	z
Acquit	1	0

with $z > 0$. The optimal action is to convict if their beliefs, β, satisfy $\beta z \geq (1-\beta)$, i.e. $\beta \geq \frac{1}{1+z}$. The signaler is the prosecuting attorney, their utility is 1 if the accused is convicted, 0 else. The signaler commits to an information structure, and following Blackwell [9, 10], this can be modeled as a dilation of p, that is, as a distribution, Q, over the space of beliefs, $\Delta(\Omega) = [0, 1]$, such that $\int_{[0,1]} \beta \, dQ(\beta) = p$. The signaler's problem is

$$\max_{Q \in \Delta(\Delta(\Omega))} Q([\tfrac{1}{1+z}, 1]) \text{ subject to } \int_{\Delta(\Omega)} \beta \, dQ(\beta) = p. \tag{27}$$

The solution is a Q that puts mass $1 - \gamma$ on 0 and γ on $\frac{1}{1+z}$, and the constraint yields $(1 - \gamma)0 + \gamma\frac{1}{1+z} = p$, i.e. $\gamma = p(1 + z)$. In equilibrium, the accused is convicted $p(1 + z) > p$ of the time.

Now suppose that the Judge/Jury has a set of possible priors, $A = [a, b]$ and we wish to solve the same design problem for the prosecuting attorney. We represent $[a, b]$ as a point in the triangle $\mathbb{K}_{\Delta(\mathbb{X})} = \{(a, b) \in [0, 1]^2 : 0 \leq a \leq b \leq 1\}$. Because $\mathbb{K}_{\Delta(\mathbb{X})}$ is a simplex, each (a, b)

has a unique representation as $(a,b) = (1-(\gamma_{0,1}+\gamma_{1,1}))(0,0) + \gamma_{0,1}(0,1) + \gamma_{1,1}(1,1)$. Solving for the weights, they are $\gamma_{1,1} = a$, $\gamma_{0,1} = (b-a)$ (the diameter of the interval of beliefs), and the weight on $(0,0)$ is $(1-b)$. Interpreting the point $(0,1) \in \mathbb{K}_{\Delta(\mathbb{X})}$ as the question of Guilt or Innocence being unknowable, the Judge/Jury's utility function can be given as

	$Innocent$	Unknowable	$Guilty$
Convict	0	$\frac{1}{2}(1-v_c)z$	z
Acquit	1	$\frac{1}{2}(1-v_a)$	0

where $v_c z > 0$ and $v_a > 0$ represent ambiguity aversion. If beliefs are given by the set $[0,1]$, then any standard of proof at least as strong as 'preponderance of evidence' makes acquital optimal, $(1-v_a) > (1-v_c)z$. We will assume something stronger, that increases in r decrease utility more if the decision is to convict than if the decision is to acquit, $0 > -v_a > -v_c z$. For beliefs $B = [p-r, p+r]$, this means that more ambiguity, a higher r, requires a higher probability of guilt, p, for conviction. The Judge/Jury convicts if their beliefs $B = [p-r, p+r]$ satisfy $U(\text{convict} \times [p-r,p+r]) \geq U(\text{acquit} \times [p-r,p+r])$, that is, if

$$pz - v_c z r \geq (1-p) - v_a r, \text{ equivalently } p \geq \tfrac{1}{1+z} + r\tfrac{v_c z - v_a}{1+z}. \tag{28}$$

As $(v_c z - v_a) > 0$, larger degrees of ambiguity, that is, larger values of r, require a larger p to convict.

Let S denote the conviction-optimal set $\left\{[a,b] \in \mathbb{K}_{\Delta(\mathbb{X})} : \frac{a+b}{2} \geq \frac{1}{1+z} + \frac{b-a}{2}\frac{v_c z - v_a}{1+z}\right\}$, so that S is the triangle bounded with vertices $(\frac{1}{1+z}, \frac{1}{1+z})$, $(1,1)$ and $(a^*,1)$ where $a^* = \frac{(1-z)-(v_c z - v_a)}{(1+z)+(v_c z - v_a)}$ (see Figure 3).

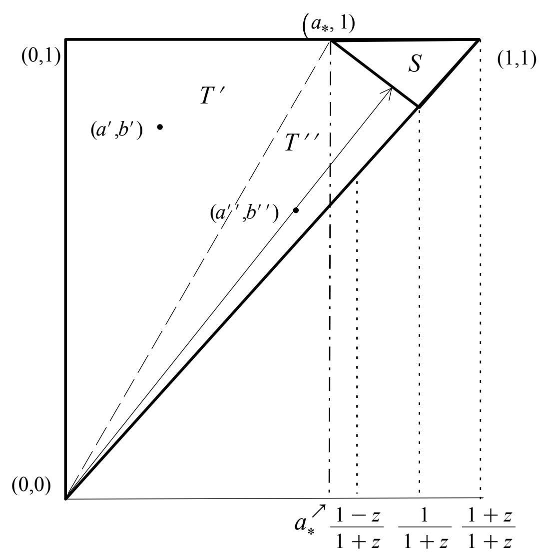


FIGURE 3

The signaler's problem is

$$\max_{Q \in \Delta(\mathbb{K}_{\Delta(\mathbb{X})})} Q(S) \text{ subject to } \int_{\mathbb{K}_{\Delta(\mathbb{X})}} B\, dQ(B) = [a,b]. \tag{29}$$

For any prior beliefs $[a',b']$ in T', the region of $\mathbb{K}_{\Delta(\mathbb{X})}$ above the line joining $(0,0)$ to $(a^*,1)$, the signaler's optimal Q puts weight only on $(0,0)$, $(0,1)$ and $(a^*,1)$ and satisfies $\gamma_{0,0}(0,0) + \gamma_{0,1}(0,1) + \gamma_{a^*,1}(a^*,1) = (a,b)$. For any initial beliefs $[a'',b'']$ in T'', the region of $\mathbb{K}_{\Delta(\mathbb{X})}$ below the line joining $(0,0)$ to $(a^*,1)$, the signaler's optimal Q puts weight only on $(0,0)$ and the first intersection of S with the ray starting at $(0,0)$ and going through (a'', b'') (the head of the arrow in the figure). In the special case that $a'' = b''$, we are in the risky case analyzed above and the signaler's optimum puts mass on $(0,0)$ and $(\frac{1}{1+z}, \frac{1}{1+z})$.

For a more general version of this model, suppose that the set of possible actions, K, is compact, as is $\mathbb{X}$. We suppose that the receiver's utility function for action $a \in K$ when they have beliefs $B \in \mathbb{K}_{\Delta(\mathbb{X})}$ is given by a jointly continuous $U(a, B)$ with each $U(a, \cdot)$ linear, and that the sender's utility function is given by a function $(a, B) \mapsto V(a, B)$ satisfying the same conditions. For $B \in \mathbb{K}_{\Delta(\mathbb{X})}$, define

$$K^*(B) = \operatorname*{argmax}_{a \in K} U(a, B), \text{ and } v(B) = \max_{a \in K^*(B)} V(a, B). \tag{30}$$

The existence of a solution to the problem in (29) is an instance of the following.

Proposition 6. *Under the compactness and continuity conditions just given, for every set of priors $A \in \mathbb{K}_{\Delta(\mathbb{X})}$, there exists a commitment-optimal signaling structure for the signaler.*

Proof. Let $A \in \mathbb{K}_{\Delta(\mathbb{X})}$ be the receiver's priors. The set $\mathbb{I} := \{Q \in \Delta(\mathbb{K}_{\Delta(\mathbb{X})}) : \int B\, dQ(B) = A\}$ is closed, hence compact. Since $K^*(\cdot)$ is upper hemicontinuous, $v(\cdot)$ is an upper semicontinuous (usc) function. This implies that the mapping $Q \mapsto \int v(B)\, dQ(B)$ is usc, hence achieves its maximum on the compact set $\mathbb{I}$. □

If the signaler is restricted to information structures where only some $E \subset \{I, G\}$ can be revealed, neither the risky nor the ambiguous version of this problem contains much of interest.

6.4. **Utility Range Dominance and Support Sets.** We now investigate how $\eta_{\min} \geq \eta_{\max}$ interacts with the support set for $(\eta_{\min} + \eta_{\max})$. The key concept is utility range dominance.

Definition 12. *For $\mathbb{V} \subset \mathbb{U}_0^1$ and $A, B \in \mathbb{K}_{\Delta(\mathbb{X})}$, the A **utility range dominates** B **on** $\mathbb{V}$, $A \succsim_{\mathbb{V}} B$, if for all $v \in \mathbb{V}$, the utility intervals have the same center,*

$$\tfrac{1}{2}[\min_{\mu \in A}\langle v, \mu\rangle + \max_{\mu \in A}\langle v, \mu\rangle] = \tfrac{1}{2}[\min_{\mu \in B}\langle v, \mu\rangle + \max_{\mu \in B}\langle v, \mu\rangle],$$

and the radius of A's utility intervals is weakly smaller,

$$\tfrac{1}{2}[\max_{\mu \in A}\langle v, \mu\rangle - \min_{\mu \in A}\langle v, \mu\rangle] \leq \tfrac{1}{2}[\max_{\mu \in B}\langle v, \mu\rangle - \min_{\mu \in B}\langle v, \mu\rangle].$$

Lemma 4. *If $(\eta_{\min} + \eta_{\max})(\mathbb{V}) = 1$, $\eta_{\min} \geq \eta_{\max}$, and $A \succsim_{\mathbb{V}} B$, then $L_\eta(A) \geq L_\eta(B)$.*

Proof. Because $\mathbb{V}$ is a carrier set for $(\eta_{\min} + \eta_{\max})$,

$$L_\eta(A) = \int_{\mathbb{V}} \min_{\mu \in A}\langle v, \mu\rangle\, d\eta_{\min}(v) + \int_{\mathbb{V}} \max_{\nu \in A}\langle v, \mu\rangle\, d\eta_{\max}(v) \tag{31}$$

with a similar expression for $L_\eta(B)$. Because $A \succsim_{\mathbb{V}} B$, for each $v \in \mathbb{V}$, the utility ranges have the same center. Therefore, there exists $r_v \geq 0$ such that $\min_{\mu \in A}\langle v, \mu\rangle = r_v + \min_{\mu \in B}\langle v, \mu\rangle$ and $\max_{\mu \in A}\langle v, \mu\rangle = (-r_v) + \max_{\mu \in B}\langle v, \mu\rangle$. This yields

$$[L_\eta(A) - L_\eta(B)] = \int_{\mathbb{V}} r_v\, d\eta_{\min}(v) + \int_{\mathbb{V}} (-r_v)\, d\eta_{\max}(v) = \tag{32}$$

$$\int_{\mathbb{V}} r_v\, d(\eta_{\min} - \eta_{\max})(v).$$

Because $r_v \geq 0$ and $\eta_{\min} \geq \eta_{\max}$, $[L_\eta(A) - L_\eta(B)] \geq 0$. □

6.4.1. *Utility Range Equality.* An implication of Lemma 4 is that $A \succsim_{\mathbb{V}} B$, $B \succsim_{\mathbb{V}} A$, and $(\eta_{\min}+\eta_{\max})(\mathbb{V}) = 1$ imply that $L_\eta(A) = L_\eta(B)$. To put this another way, L_η ignores expansions in 'directions' not included in the support set of $(\eta_{\min} + \eta_{\max})$.

Example 13. *Let $\mathbb{X} = \{a, b, c\}$ with $a \prec b \prec c$, so that $\mathbb{ND} \cap \mathbb{U}_0^1$ is the set of utility functions $u_r = (u_r(a), u_r(b), u_r(c)) = (0, r, 0)$, $r \in [0, 1]$. Let $\mu_1 = (0, 0.3, 0.7)$, $\mu_2 = (0.3, 0.3, 0.4)$, $\mu_3 = (0, 0.6, 0.4)$, and $\mu_4 = (0.3, 0, 0.7)$ in $\Delta(\mathbb{X})$ so that $\mu_1 \succsim_F \mu_2, \mu_3, \mu_4$ and $\mu_3, \mu_4 \succsim_F \mu_2$. Figure 3 illustrates the sets $C = \{\mu \in \Delta(\mathbb{X}) : \mu_1 \succsim_F \mu \succsim_F \mu_2\}$***co**$(\mu_1, \mu_2, \mu_3, \mu_4)$, $A =$ **co**(μ_1, μ_2), and a set $B \in \mathbb{K}_{\Delta(\mathbb{X})}$ with $A \subset B \subset C$. For any η supported on $\mathbb{ND} \subset \mathbb{U}_0^1$, $L_\eta(A) = L_\eta(B) = L_\eta(C)$*

— for any u_r, μ_1 is a best element in A, B, and C while μ_2 is a worst element. These mean that the utility intervals $[\inf_{\mu\in A}, \sup_{\mu\in A}]$, $[\inf_{\mu\in B}, \sup_{\mu\in B}]$, *and* $[\inf_{\mu\in C}, \sup_{\mu\in C}]$ *are equal.*

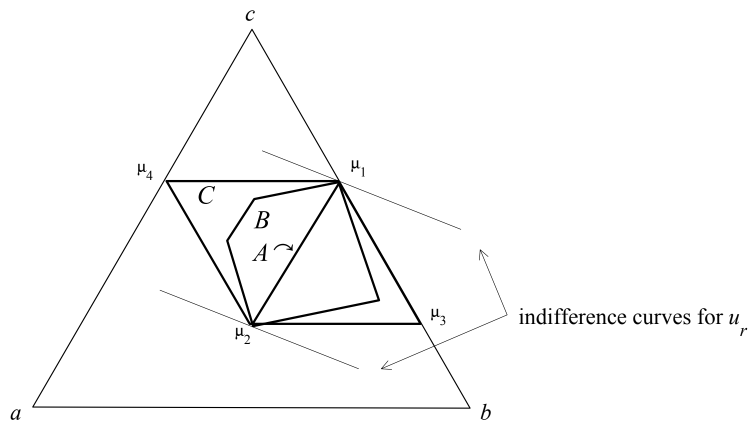


FIGURE 4

6.4.2. *Utility Range Dominance and Sources of Uncertainty.* Machina [39] shows that preferences over random variables satisfying the rank-dependence axiom must be indifferent over the sets in the following example.

Example 14. *Suppose that we toss a fair coin indepedently of drawing a ball from an urn known to contain balls that are either Black or White with no information about the proportions. Consider the two sets of probabilities over the three outcomes, -8, 0, and $+8$ given by the following tables.*

A	*Black*	*White*
Heads	+8	0
Tails	−8	0

B	*Black*	*White*
Heads	0	0
Tails	−8	+8

$u_r \in \mathbb{ND} \cap \mathbb{U}_0^1$ *iff it is of the form* $u_r = (u_r(-8), u_r(0), u_r(+8)) = (0, r, 1)$ *for some* $r \in [0, 1]$. *Figure* 5(*a*) *illustrates the sets of probabilities A and B, and Figure* 5(*b*) *illustrates, as a function of r, the utility intervals*

$$U_{B,r} := [\min_{\mu\in B}\langle u_r, \mu\rangle, \max_{\mu\in B}\langle u_r, \mu\rangle] \text{ and} \tag{33}$$

$$U_{A,r} := [\min_{\mu\in A}\langle u_r, \mu\rangle, \max_{\mu\in A}\langle u_r, \mu\rangle]. \tag{34}$$

For each r, these intervals have the same center, and for $r \neq 0, 1$, the set $U_{B,r}$ is strictly wider than $U_{A,r}$. If $(\eta_{\min} + \eta_{\max})(\mathbb{ND}) = 1$ and $\eta_{\min} \geq \eta_{\max}$, then $L_\eta(A) \geq L_\eta(B)$ with strict inequality if $\eta_{\min} > \eta_{\max}$ and $(\eta_{\min} + \eta_{\max})$ puts positive weight on the set of u_r with $0 < r < 1$.

Comments.

a. The example works because, across the set A, possible expected utility levels do not vary as much as they do across the set B. This is where the difference in the sources of uncertainty enters. In the set A, the larger utility range uncertainty, -8 to $+8$, is determined by a known probability, $\frac{1}{2}$, while in the set B the larger range utility uncertainty is determined by an unknown probability. As Machina [39] shows, the symmetry built into models of preferences over random variables that satisfy a rank-dependence axiom *requires* indifference between the sets A and B.

b. If the coin is "bent," that is, slightly biased in an unknown direction, the set A becomes the triangular set

$$A_b = \mathbf{co}\left(\{(\tfrac{1}{2} - b, 0, \tfrac{1}{2} + b), (\tfrac{1}{2} + b, 0, \tfrac{1}{2} - b), (0, 1, 0)\}\right),$$

i.e. becomes slightly flared when it meets the $(\alpha, 0, (1-\alpha))$ face of the vertex, while the set B becomes

$$B_b = \mathbf{co}\,(\{(0, \tfrac{1}{2} - b, \tfrac{1}{2} + b), (0, \tfrac{1}{2} + b, \tfrac{1}{2} - b), (\tfrac{1}{2} - b, \tfrac{1}{2} + b, 0), (\tfrac{1}{2} + b, \tfrac{1}{2} - b, 0)\}),$$

i.e. becomes a quadrilateral having slightly wider intersections with the $(\beta, (1-\beta), 0)$ and the $(0, \gamma, (1-\gamma))$ faces. By continuity, if $L_\eta(A) > L_\eta(B)$, then $L_\eta(A_b) > L_\eta(B_b)$ for small biases b.

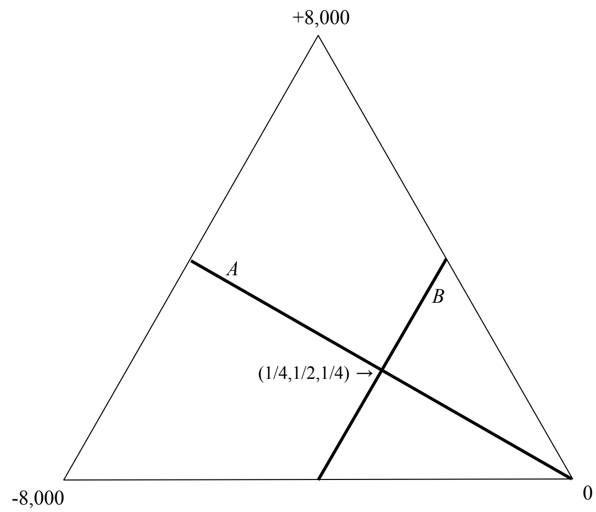


Figure 5(a)
$\mathrm{St}(A) = \mathrm{St}(B) = (1/4, 1/2, 1/4)$

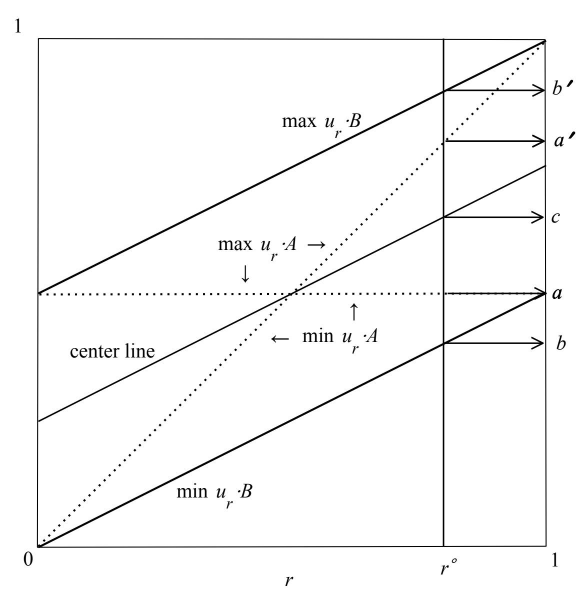


Figure 5(b)
$[a, a'] = [\min u_{r^\circ} \cdot A, \max u_{r^\circ} \cdot A]$
$[b, b'] = [\min u_{r^\circ} \cdot B, \max u_{r^\circ} \cdot B]$
$c = \frac{1}{2}a + \frac{1}{2}a' = \frac{1}{2}b + \frac{1}{2}b'$

6.4.3. *The Ambiguity of "Crisp" Acts.* Following [25] (and a slightly stronger definition from [14, Theorem 12]), a sufficient condition for a random variable $f : \Omega \to \mathbb{X}$ to be **crisp** relative to a set of priors S is that $f(S) \subset \boldsymbol{u}^{-1}(r)$ for some $r \in \mathbb{R}$, equivalently, if for all $p \in S$, $\int \boldsymbol{u}(f(\omega))\, dp(\omega) = r$. Such acts can still be ambiguous.

Example 15. *Consider the preferences on $\mathbb{K}_{\Delta(\mathbb{X})}$ when $\mathbb{X} = \{-8,000, 0, +8,000\}$ represented by L_η where $\eta_{\min}$ and $\eta_{\max}$ are carried by the $u_r \in \mathbb{ND} \cap \mathbb{U}_0^1$ of the form $u_r = (u_r(-8), u_r(0), u_r(+8)) = (0, r, 1)$ for some $r \in [0,1]$, and suppose that $\eta_{\min} = \frac{2}{3}\lambda$ and $\eta_{\max} = \frac{1}{3}\lambda$, λ being the uniform distribution on $[0,1]$, so that $\eta_{\min} > \eta_{\max}$. The resultant, $\boldsymbol{u}$, is the risk neutral vNM utility function $u_{\frac{1}{2}}$ so that, provided the decision maker can conceive of them, any subset of the set A given in Figure 5(a) can arise from a crisp act. If $C = \{\mu\}$ is a singleton subset of A, then $L_\eta(C) = \int_{\mathbb{X}} \boldsymbol{u}(x)\, d\mu(x) = \frac{1}{2}$. By contrast, for the ambiguous set A, $L_\eta(A) = \frac{11}{24} < \frac{1}{2}$.*

We have already seen in Example 4 that smooth versions of MBA preferences must be ambiguity neutral in a neighborhood of the set of risky outcomes. What the previous example shows is that they must also be ambiguity neutral in the neighborhood of a class of ambiguous outcomes.

6.5. **Smooth Models of Ambiguity Aversion.** The utility range analysis just given captures part of the appeal of the smooth models of ambiguity aversion developed Klibanoff *et al.* These give the utility of a function $f : \Omega \to \mathbb{X}$ as

$$\text{Smooth}(f) = \int_\Pi \varphi(\langle \boldsymbol{u}, f(p)\rangle)\, dQ(p) \tag{35}$$

where $\boldsymbol{u} \in \mathbb{U}_0^1$, $\varphi : [0,1] \to [0,1]$ is concave and Q is a probability distribution on a set of priors Π.

Unlike the previously discussed preferences, change of variables does not allow smooth preferences to fit into the framework of preferences over $\mathbb{K}_{\Delta(\mathbb{X})}$. Rather, the domain becomes $\{(A, \mathfrak{p}) : A \in \mathbb{K}_{\Delta(\mathbb{X})}, \mathfrak{p} \in \Delta(A)\}$ and $\text{Smooth}_{cov}(A, \mathfrak{p})$ is

$$\text{Smooth}_{cov}(A, \mathfrak{p}) = \int_A \varphi(\langle \boldsymbol{u}, \mu\rangle)\, d\mathfrak{p}(\mu). \tag{36}$$

As above, if Π is not descriptively complete, a DM modeled using preferences over measurable functions will only be able to conceive of a limited subset of the domain of $\text{Smooth}_{cov}(\cdot)$. The next result is a direct consequence of Theorem 2 and the Blackwell and Dubins [11] extension of the Skorohod representation theorem.

Lemma 5. *If S is a descriptively complete set of priors and Q is a non-atomic distribution on S, then for every $(A, \mathfrak{p})$, there exists an $f : \Omega \to \mathbb{X}$ such that $\mathfrak{p}$ is the distribution of $f(p)$ when $p \in S$ is distributed according to Q.*

7. Summary and Future Directions

This paper has developed two concepts: the representation theory for continuous linear functionals on compact convex sets of probability distributions over utility relevant outcomes; the theory of descriptively complete sets of priors. The representation theory leads to much more tractable modeling of ambiguous choice problems. The theory of descriptively complete sets of priors resolves several open problems and allows the multiple prior preferences over measurable functions approach to model as large a class of problems as the continuous linear preferences presented here.

In applied theory models of risky choice situations, the modeler must specify the decision maker's vNM utility function on $\Delta(\mathbb{X})$ and what distribution that the decision maker believes to be associated with any given choice. The theory of choice under ambiguity developed here requires the same level of detail, the modeler must specify the decision maker's vNM utility function on $\mathbb{K}_{\Delta(\mathbb{X})}$ and what *set* of distributions the decision maker believes to be associated with any given choice. The essential reason that the von Neumann/Morgenstern approach to ambiguous decision problems is more tractable than the random variable approach is that it starts with a specification of the risk, the ambiguity, and the attitudes toward them rather than trying to derive all of these from preferences over a space of measurable functions.

Descriptive completeness of a set of priors is necessary for the modeled decision maker to be able to conceive of every element in some non-empty open set of partial descriptions. Given descriptive completeness and state independence, the preferences over measurable functions approach to the theory of choice in the presence of ambiguity can be understood as the study of preferences over sets of probability distributions. This change of perspective allows us to develop, in particular, the theory of comparative ambiguity aversion much further than has been possible with multiple prior models.

It seems to us that there are many classes of applications where the introduction of ambiguity will make a substantial difference, that there are some fascinating interpretational issues yet to be resolved, and that there are some issues of a more theoretical nature as well. We discuss these in turn.

7.1. **Possible Applications.** There is clearly a great scope for extensions of monotone comparative statics results to ambiguous problems. The partial orders given by first and second order dominance, as well as utility range dominance should play a central role. This will require a systematic investigation of how other classes of partial orders on probabilities extend to ambiguous problems.

One of our initial motivations for studying this class of problems was Knight's [35] observation that most R&D takes place in larger firms because they can afford several projects simultaneously. His intuition was that even though each such project would have an unknown distribution over rewards, some version of the strong law of large numbers/central limit theorem should imply that several projects should yield less relative ambiguity. This seems to us to be the same intuition that is behind Schmeidler's [48] definition of ambiguity aversion as a kind of preference for diversification. This will require a systematic investigation of how convolution of sets of distributions interacts with ambiguity aversion and decreasing relative ambiguity aversion.

For the class of problems $\mathbb{K}^{sym}_{\Delta(\mathbb{X})}$ and $\mathbb{K}^{fin}_{\Delta(\mathbb{X})}$, we can completely separate attitudes toward risk and attitudes toward ambiguity. This should allow elicitation of the tradeoffs between the two. A particularly useful form of elicitation in risky problems are the proper scoring rules. This, and generic control of the elicitee's value function [23] should be possible for ambiguous problems as well.

One of the criticisms of mechanism design is the sensitive dependence of optimal contracts on details of the model. An alternative approach to robust mechanism design may be to assume that the principal's prior over the agent's utility and information is a set rather than a point. We suspect that optimal contracts will need to take into account more of the possibilities of error, and this may lead to a more 'central' and stable class of optimal contracts.

7.2. **Interpretational Issues.** Information structures for risky problems are modeled as distributions over posterior distributions having the prior as the resultant. For ambiguous problems this same structure is available with one change — we replace posteriors by sets of posteriors and replace the single prior with a single set of priors. Of particular interest are the best information structures, the distributions carried on the extreme points of $\Delta(\mathbb{X})$ in the risky case, and the distributions carried on the extreme sets of $\mathbb{K}_{\Delta(\mathbb{X})}$ in the ambiguous case. Formally, the extreme sets in $\mathbb{K}_{\Delta(\mathbb{X})}$ can be treated as if they are a new class of points. For example, the table giving the Judge/Jury's utility function includes the "Unknowable" state. A prosaic interpretation of such states might run along the lines of "unknowable in the time frame available for making a decision." A more speculative interpretation might, in some problems, run along the lines of previously unobserved states, of states not available in the original model.

Perhaps the most fascinating extant interpretation of decision makers using the multiple prior preferences is that they are reasoning by analogy [3]. It seems to us that this may offer another way to understand descriptive completeness of a set of priors. In particular, we would like to know what structure or structures of the set of analogies is sufficient, or necessary, for the implied set of priors to be descriptively complete.

On a different topic, writing down the conditions for an ambiguous Nash equilibrium in sets of distributions is immediate. However, one often interprets only having a partial description of what is happening as a lack of experience with the problem. By contrast, equilibrium is often understood as what happens when there is sufficient experience. It may be that ambiguous equilibria will, at least some of the time, describe intermediate-run behavior along some learning dynamics, a result that would greatly clarify their interpretation. It may also be possible to purify ambiguous equilibra using sets of perturbations to utilities, and this would provide another class of interpretations.

7.3. **Extensions to the Theory.** Skorohod's representation theorem has two parts: representation, any μ is of the form $f(p)$ provide that $p \in \Delta(\Omega)$ is non-atomic; and continuity, if $\mu_n \to_w \mu$, then there is a sequence, $f, (f_n)_{n\in\mathbb{N}}$, such that $f_n(p) = \mu_n$, $f(p) = \mu$, and $f_n(\omega) \to f(\omega)$ for a set of ω having p-mass 1. This paper has only used the representation part of the set-valued Skorohod theorem in [19], the continuity part is, so far, absolutely unexploited. We suspect that appropriate versions of the Theorem of the Maximum and the various upper hemicontinuity results in game theory and other equilibrium models will be relatively easy to prove using almost everywhere convergence. The representation part of the result extends to closed non-compact sets of probabilities, indeed, it extends to measurable sets of probabilities. As can be seen in [8], in the non-compact case, the appropriate definition of continuity is a far more subtle issue.

The separations of ambiguity attitude and risk attitude depend on a continuous direct sum decomposition for subclasses of $\mathbb{K}_{\Delta(\mathbb{X})}$. A Minkowski class of sets is one that is closed under addition and non-negative dilation, and these are the basic ingredients for linear functionals on sets and ambiguity aversion respectively. It may be possible to extend the separation of attitudes toward risk and ambiguity to larger Minkowski classes of sets in [50].

Understanding of non-linear preferences over Minkowski subclasses of $\mathbb{K}_{\Delta(\mathbb{X})}$ may also benefit from study of their local linear approximations. For finite $\mathbb{X}$, let $\mathbb{K}^{full}_{\Delta(\mathbb{X})} \subset \mathbb{K}_{\Delta(\mathbb{X})}$ denote the full dimensional subsets of $\mathbb{K}_{\Delta(\mathbb{X})}$, that is, the set of A with $\dim(A) = \dim(\Delta(\mathbb{X}))$. The preferences on $\mathbb{K}^{full}_{\Delta(\mathbb{X})}$ in [1] can be represented by the utility function $V(A) = E^Q(f|A)$ where $f : \Delta(\mathbb{X}) \to \mathbb{R}$ is continuous and $Q \in \Delta(\Delta(\mathbb{X}))$ has an almost everywhere positive density with respect to Hausdorff/Lebesgue measure on $\Delta(\mathbb{X})$. If f is not constant, then $V(\cdot)$ is not Lipschitz and does not have a continuous extension from the dense class $\mathbb{K}^{full}_{\Delta(\mathbb{X})}$ to $\mathbb{K}_{\Delta(\mathbb{X})}$ when $\#\mathbb{X} \geq 3$. However, restricted to $\mathbb{K}^{full}_{\Delta(\mathbb{X})}$, $V(\cdot)$ is locally Lipschitz at most A if the density of Q is well-behaved. In Banach spaces, Lipschitz functions are differentiable outside of Gauss null sets, so it may be that there are informative local linear approximations to $V(\cdot)$ at most points in its domain.

Theorem 3 showed that, when $\mathbb{X} = [0, M]$, descriptive completeness of a set of priors is equivalent to the decision maker being able to contemplate a single non-empty open subset of $\mathbb{K}_{\Delta(\mathbb{X})}$. This property leads us to conjecture that for generic pairs of convex sets of priors that fail to be descriptively complete, the overlap in the problems that they can model is quite minimal. As a side benefit to resolving this question, it may be possible to order sets of priors in a more continuous fashion on the basis of the class of problems they allow one to model.

References

[1] David S. Ahn. Ambiguity without a state space. *The Review of Economic Studies*, 75(1):3–28, 2008.

[2] Massimiliano Amarante. Foundations of neo-Bayesian statistics. *J. Econom. Theory*, 144(5):2146–2173, 2009.

[3] Massimiliano Amarante. Analogy in decision-making. *Working paper, Department of Economics, Université de Montréal et CIREQ*, 2012.

[4] Robert M. Anderson and William R. Zame. Genericity with infinitely many parameters. *Adv. Theor. Econ.*, 1:Art. 1, 64 pp. (electronic), 2001.

[5] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Math. Finance*, 9(3):203–228, 1999.
[6] Susan Athey. Monotone comparative statics under uncertainty. *The Quarterly Journal of Economics*, 117(1):187–223, 2002.
[7] Heinz Bauer. Kennzeichnung kompakter Simplexe mit abgeschlossener Extremalpunktmenge. *Arch. Math.*, 14:415–421, 1963.
[8] Gerald Beer. *Topologies on Closed and Closed Convex Sets*, volume 268 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1993.
[9] David Blackwell. Comparison of experiments. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*, pages 93–102, Berkeley and Los Angeles, 1951. University of California Press.
[10] David Blackwell. Equivalent comparisons of experiments. *Ann. Math. Statistics*, 24:265–272, 1953.
[11] David Blackwell and Lester E. Dubins. An extension of Skorohod's almost sure representation theorem. *Proc. Amer. Math. Soc.*, 89(4):691–692, 1983.
[12] Ethan D. Bolker. A class of convex bodies. *Trans. Amer. Math. Soc.*, 145:323–345, 1969.
[13] Leo Breiman, Lucien Le Cam, and Lorraine Schwartz. Consistent estimates and zero-one sets. *Ann. Math. Statist.*, 35:157–161, 1964.
[14] Simone Cerreia-Vioglio, Paolo Ghirardato, Fabio Maccheroni, Massimo Marinacci, and Marciano Siniscalchi. Rational preferences under ambiguity. *Econom. Theory*, 48(2-3):341–375, 2011.
[15] Soo Hong Chew, Larry G. Epstein, and Itzhak Zilcha. A correspondence theorem between expected utility and smooth utility. *J. Econom. Theory*, 46(1):186–193, 1988.
[16] Vincent P. Crawford and Joel Sobel. Strategic information transmission. *Econometrica*, 50(6):1431–1451, 1982.
[17] Eddie Dekel. An axiomatic characterization of preferences under uncertainty: weakening the independence axiom. *J. Econom. Theory*, 40(2):304–318, 1986.
[18] Claude Dellacherie and Paul-André Meyer. *Probabilities and Potential*, volume 29 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 1978.
[19] Martin Dumav and Maxwell Stinchcombe. Skorohod's representation theorem for sets of probabilities. *Working paper, Department of Economics, University of Texas at Austin*, 2011.
[20] Larry Epstein and Shaolin Ji. Ambiguous volatility, possibility and utility in continuous time. *arXiv preprint arXiv:1103.1652v6*, 2012.
[21] Larry G. Epstein. A definition of uncertainty aversion. *Rev. Econom. Stud.*, 66(3):579–608, 1999.
[22] Larry G. Epstein and Jiankang Zhang. Subjective probabilities on subjectively unambiguous events. *Econometrica*, 69(2):265–306, 2001.
[23] Fang Fang, Maxwell B. Stinchcombe, and Andrew B. Whinston. Proper scoring rules with arbitrary value functions. *J. Math. Econom.*, 46(6):1200–1210, 2010.
[24] Thibault Gajdos, Takashi Hayashi, Jean-Marc Tallon, and Jean-Christophe Vergnaud. Attitude toward imprecise information. *J. Econom. Theory*, 140(1):27–65, 2008.
[25] Paolo Ghirardato, Fabio Maccheroni, and Massimo Marinacci. Differentiating ambiguity and ambiguity attitude. *J. Econom. Theory*, 118(2):133–173, 2004.
[26] Itzhak Gilboa and David Schmeidler. Maxmin expected utility with nonunique prior. *J. Math. Econom.*, 18(2):141–153, 1989.
[27] Sander C. Hille and Daniël T. H. Worm. Embedding of semigroups of Lipschitz maps into positive linear semigroups on ordered Banach spaces generated by measures. *Integral Equations Operator Theory*, 63(3):351–371, 2009.
[28] Lars Hörmander. Sur la fonction d'appui des ensembles convexes dans un espace localement convexe. *Arkiv för Matematik*, 3:181–186, 1955.
[29] Jean-Yves Jaffray. Généralisation du critère de l'utilité espérée aux choix dans l'incertain régulier. *Recherche opérationnelle*, 23:237–267, 1989.
[30] Jean-Yves Jaffray. Linear utility theory for belief functions. *Operations Research Letters*, 8(2):107–112, 1989.
[31] Jean-Yves Jaffray and Fabrice Philippe. On the existence of subjective upper and lower probabilities. *Math. Oper. Res.*, 22(1):165–185, 1997.
[32] Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *The American Economic Review*, 101(6):2590–2615, 2011.
[33] Peter Klibanoff, Massimo Marinacci, and Sujoy Mukerji. A smooth model of decision making under ambiguity. *Econometrica*, 73(6):1849–1892, 2005.
[34] Peter Klibanoff, Sujoy Mukerji, and Kyoungwon Seo. Relevance and symmetry. *Working paper, Department of Economics, Northwestern University*, 2012.
[35] Frank H. Knight. *Risk, Uncertainty and Profit*. Houghton Mifflin Company, Boston and New York, 1921.
[36] Fabio Maccheroni, Massimo Marinacci, and Aldo Rustichini. Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, 74(6):1447–1498, 2006.
[37] Mark J. Machina. "Expected Utility" Analysis without the Independence Axiom. *Econometrica*, 50(2):277–323, 1982.

[38] Mark J. Machina. Risk, ambiguity, and the rank-dependence axioms. *The American Economic Review*, 99(1):385–392, 2009.
[39] Mark J. Machina. Ambiguity aversion with three or more outcomes. *Working paper, Department of Economics, U. C. San Diego*, 2012.
[40] Ernest Michael. Continuous selections. I. *Ann. of Math. (2)*, 63:361–382, 1956.
[41] Ilya S. Molchanov. *Limit Theorems for Unions of Random Closed Sets*, volume 1561 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1993.
[42] W. Olszewski. Preferences over sets of lotteries. *The Review of Economic Studies*, 74(2):567–595, 2007.
[43] K. R. Parthasarathy. *Probability Measures on Metric Spaces*. AMS Chelsea Publishing, Providence, RI, 2005. Reprint of the 1967 original.
[44] Robert R. Phelps. Infinite dimensional compact convex polytopes. *Math. Scand.*, 24:5–26, 1969.
[45] Robert R. Phelps. *Lectures on Choquet's Theorem*, volume 1757 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, second edition, 2001.
[46] Leonard J. Savage. *The Foundations of Statistics*. Dover Publications Inc., New York, revised edition, 1972.
[47] Alexander Schied. Risk measures and robust optimization problems. *Stoch. Models*, 22(4):753–831, 2006.
[48] David Schmeidler. Subjective probability and expected utility without additivity. *Econometrica*, 57(3):571–587, 1989.
[49] Rolf Schneider. *Convex Bodies: the Brunn-Minkowski Theory*, volume 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1993.
[50] Rolf Schneider and Franz E. Schuster. Rotation invariant Minkowski classes of convex bodies. *Mathematika*, 54(1-2):1–13, 2007.
[51] Marciano Siniscalchi. A behavioral characterization of plausible priors. *J. Econom. Theory*, 128(1):91–135, 2006.
[52] Anatoliĭ V. Skorohod. Limit theorems for stochastic processes. *Teor. Veroyatnost. i Primenen.*, 1:289–319, 1956.
[53] George J. Stigler and Gary S. Becker. De gustibus non est disputandum. *The American Economic Review*, 67(2):76–90, 1977.
[54] Maxwell B. Stinchcombe. Countably additive subjective probabilities. *Rev. Econom. Stud.*, 64(1):125–146, 1997.
[55] Richard A. Vitale. The Steiner point in infinite dimensions. *Israel J. Math.*, 52(3):245–250, 1985.
[56] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, anniversary edition, 2007. With an introduction by Harold W. Kuhn and an afterword by Ariel Rubinstein.
[57] Peter P. Wakker. Testing and characterizing properties of nonadditive measures through violations of the sure-thing principle. *Econometrica*, 69(4):1039–1059, 2001.
[58] Nik Weaver. *Lipschitz Algebras*. World Scientific Publishing Co. Inc., River Edge, NJ, 1999.

Appendix A. The Proof of the Representation Theorem for Polish Spaces

The following encapsulates the mathematical structure of vNM preferences in the theory of choice under uncertainty: the objects over which people have linear preferences are a convex subset, $\mathcal{C}$, of a Banach space, $\mathcal{X}$; the Banach space has a pre-dual, $\mathcal{X}_\circ$; the norms on $\mathcal{X}$ and $\mathcal{X}_\circ$ are related by $\|x\| = \sup_{\|x_\circ\|\le 1} |\langle x, x_\circ\rangle|$ and $\|x_\circ\|_\circ = \sup_{\|x\|\le 1} |\langle x, x_\circ\rangle|$; the relevant set of linear preferences are given by $U_{x_\circ}(x) = \langle x, x_\circ\rangle$ where $x_\circ \in \mathcal{X}_\circ$; these linear functionals are continuous in the weak* topology, defined by $x_n \to_{w^*} x$ in $\mathcal{C}$ iff $\langle x_n, x_\circ\rangle \to \langle x, x_\circ\rangle$ for all $x_\circ \in \mathcal{X}_\circ$. One then studies properties of expected utility preferences by studying properties of the elements in the pre-dual that define the continuous linear functionals.

1. For expected utility preferences with a compact set of consequences: $\mathcal{C}$ is the set of countably additive probabilities, $\Delta(\mathbb{X})$; $\mathcal{X} = \mathbf{ca}(\mathbb{X})$ is the set of countably additive, finite signed measures with the variation norm; $\mathcal{X}_\circ = C(\mathbb{X})$; the set of continuous linear functions are of the form $U(p) = \langle u, p\rangle = \int_{\mathbb{X}} u(x)\, dp(x)$ where $u \in C(\mathbb{X})$; and $p_n \to_{w^*} p$ in $\Delta(\mathbb{X})$ iff for every $u \in C(\mathbb{X})$, $\langle u, p_n - p\rangle = \int_{\mathbb{X}} u(x)\, d(p_n - p)(x) \to 0$.
2. For expected utility theory with a general Polish set of consequences: $\mathcal{X}_\circ = C_b(\mathbb{X})$; $\mathcal{X}$ is the set of finitely additive, finite signed measures with the variation norm; the set of continuous linear functions are of the form $U(p) = \langle u, p\rangle = \int_{\mathbb{X}} u(x)\, dp(x)$ where $u \in C_b(\mathbb{X})$; and $p_n \to_{w^*} p$ in $\Delta(\mathbb{X})$ iff for every $u \in C_b(\mathbb{X})$, $\langle u, p_n - p\rangle = \int_{\mathbb{X}} u(x)\, d(p_n - p)(x) \to 0$. Because the failure of countable additivity leads to money pumps in model with countably many consequences,

and to indifference between a random variable act and one that pointwise dominates it, we restrict attention to countably additive measures and probabilities.[13]

As above, $\mathbb{U}_0^1 := \{f \in C_b(\mathbb{X}) : \inf_{x\in\mathbb{X}} f(x) = 0, \sup_{x\in\mathbb{X}} f(x) = 1\}$ so that every element $g \in C_b(\mathbb{X})$ has a unique representation of the form $r \cdot h + s$ where $r \in \mathbb{R}_+$, $s \in \mathbb{R}$, and $h \in \mathbb{U}_0^1$. $\mathrm{Lip}(\mathbb{U}_0^1)$ denotes the set of Lipschitz functionals on $\mathbb{U}_0^1$ with the norm $\|\psi\|_{\mathrm{Lip}} = \sup_{f\in\mathbb{U}_0^1} |\psi(f)| + \sup_{f\neq g\in\mathbb{U}_0^1} \frac{|\psi(f)-\psi(g)|}{\|f-g\|_\infty}$. From the monograph/textbook [58] or [27, Cor. 3.10], $\mathcal{M}_s(\mathbb{U}_0^1)$, the set of separably supported, countably additive, finite signed measures on $\mathbb{U}_0^1$ is a pre-dual of $\mathrm{Lip}(\mathbb{U}_0^1)$.

With these notations in place, we can give the parallel mathematical structure for continuous linear preferences on sets of probabilities.

3. For preferences in the presence of ambiguity with a general Polish set of consequences: $\mathcal{C}$ is $\mathbb{K}_{\Delta(\mathbb{X})}$, the set of non-empty, compact, convex subsets of $\Delta(\mathbb{X})$; each $A \in \mathbb{K}_{\Delta(\mathbb{X})}$ is identified with its support function, $\mu_A(f) := \min_{p\in A}\langle f, p\rangle$, which is an element of $\mathcal{X} = \mathrm{Lip}(\mathbb{U}_0^1)$; the predual, $\mathcal{X}_\circ$, is the closure in $\mathcal{X}^*$ of $\mathcal{M}_s(\mathbb{U}_0^1)$, the set of separably supported, countably additive, finite signed measures on $\mathbb{U}_0^1$; the set of continuous linear functionals is given by $U(A) = \int_{\mathbb{U}_0^1} \mu_A(f)\, d\eta(f)$ where $\eta \in \mathcal{M}_s(\mathbb{U}_0^1)$; and $A_n \to_{w^*} A$ iff for every $\eta \in \mathcal{M}_s(\mathbb{U}_0^1)$, $\int_{\mathbb{U}_0^1} \mu_{A_n}(f)\, d\eta(f) \to \int_{\mathbb{U}_0^1} \mu_A(f)\, d\eta(f)$.

The following shows that weak*-convergence of A_n to A in $\mathbb{K}_{\Delta(\mathbb{X})}$ has exactly the right form.

Lemma 6. *The following are equivalent:*
(a) $d_H(A_n, A) \to 0$,
(b) *for all* $f \in \mathbb{U}_0^1$, $\mu_{A_n}(f) \to \mu_A(f)$, *and*
(c) *for all* $\eta \in \mathcal{M}_s(\mathbb{U}_0^1)$, $L_\eta(A_n) \to L_\eta(A)$.

The equivalence of (b) and (c) is a direct consequence of [58, Theorem 2.2.2].

Proof. (a)⇒(b): If $d_H(A_n, A) \to 0$, then for all $f \in \mathbb{U}_0^1$, $\min_{p\in A_n}\langle f, p\rangle \to \min_{p\in A}\langle f, p\rangle$ by the Theorem of the Maximum, proving (b).

(b)⇒(c): If $\mu_{A_n}(f) \to \mu_A(f)$ for all $f \in \mathbb{U}_0^1$, then dominated convergence implies that for any $\eta \in \mathcal{M}_s(\mathbb{U}_0^1)$, $\int \min_{p\in A_n}\langle f, p\rangle\, d\eta(f) \to \int \min_{p\in A}\langle f, p\rangle\, d\eta(f)$.

(c)⇒(a): Suppose that for all $\eta \in \mathcal{M}_s(\mathbb{U}_0^1)$, $L_\eta(A_n) \to L_\eta(A)$. Since the span of $\mathbb{U}_0^1$ and $\{1\}$ is equal to $C_b(\mathbb{X})$, from [43, Theorem 6.6], there exist a countably $\{f_k : k \in \mathbb{N}\} \subset \mathbb{U}_0^1$ such that $r(\mu, \mu') = \sum_k \frac{1}{2^k}|\langle f_k, \mu - \mu'\rangle|$ metrizes weak* convergence. Because we are considering the Hausdorff convergence of compact subsets of $\Delta(\mathbb{X})$, any metric for the weak* topology can be used ([8, Ch. 3]). Because $r(\mu, \mu') = 0$ iff $\mu = \mu'$, every $B \in \mathbb{K}_{\Delta(\mathbb{X})}$ satisfies

$$B = \cap_k \{q \in \Delta(\mathbb{X}) : \min_{p\in B}\langle p, f_k\rangle \le \langle q, f_k\rangle \le \max_{p\in B}\langle p, f_k\rangle\}.$$

Therefore, $d_H(A_n, A) \to 0$ iff for all f_k, $\min_{p\in A_n}\langle p, f_k\rangle \to \min_{p\in A}\langle p, f_k\rangle$ and $\max_{p\in A_n}\langle p, f_k\rangle \to \max_{p\in A}\langle p, f_k\rangle$. Let $\eta_1 = \sum_k \frac{1}{2^k}\delta_{f_k}$ and $\eta_2 = \sum_k \frac{1}{2^k}\delta_{-f_k}$. Because η_1 is countably supported, $L_{\eta_1}(A_n) \to L_\eta(A)$ iff for all f_k, $\min_{p\in A_n}\langle p, f_k\rangle \to \min_{p\in A}\langle p, f_k\rangle$. Because $\min_{p\in A}\langle -f, p\rangle = -\max_{p\in A}\langle f, p\rangle$, $L_{\eta_2}(A_n) \to L_\eta(A)$ iff for all f_k, $\max_{p\in A_n}\langle p, f_k\rangle \to \max_{p\in A}\langle p, f_k\rangle$. □

Recall that $L_{\eta^\circ}(A) := \int_{\mathbb{U}_0^1} \min_{\mu\in A}\langle u, \mu\rangle\, d\eta^\circ(u)$, $\mathcal{L}^\circ := \{L_{\eta^\circ} : \eta^\circ \in \mathcal{M}^s(\mathbb{U}_0^1)\}$, and $\mathcal{L} := \mathrm{cl}(\mathcal{L}^\circ)$ in the metric $d(L, L') = \sup_{A\in\mathbb{K}_{\Delta(\mathbb{X})}} |L(A) - L'(A)|$. We are now in a position to prove the representation theorem.

Theorem 1. $L : \mathbb{K}_{\Delta(\mathbb{X})} \to \mathbb{R}$ is continuous and linear if and only if $L \in \mathcal{L}$.

Proof. Because the diameter of $\mathbb{U}_0^1$ is 1, [27, Cor. 3.10] implies that $\mathcal{M}_s(\mathbb{U}_0^1)$ is $\|\cdot\|^*_{\mathrm{Lip}}$-dense in the dual of $\mathrm{Lip}(\mathbb{U}_0^1)$, and $d(L, L') = \sup_{A\in\mathbb{K}_{\Delta(\mathbb{X})}} |L(A) - L'(A)|$ metrizes this dual space norm. □

For conditions guaranteeing that $\mathcal{L}^\circ$ is not merely dense in but is equal to $\mathcal{L}$, see [27, Thm. 3.11].

[13] A collection of, and common resolution to these paradoxes, and to any others that might arise from failures of countable additivity in decision theory can be found in [54].

One clarification may be in order. In the expected utility case, the vector space containing $\Delta(\mathbb{X})$ was its span. By contrast, the span of the support functions is a strict subset of the vector space $\mathrm{Lip}(\mathbb{U}_0^1)$. Let $\mathcal{S} = \mathrm{span}(\{\mu_A : A \in \mathbb{K}_{\Delta(\mathbb{X})}\})$ denote the span of the support functions in $\mathrm{Lip}(\mathbb{U}_0^1)$. One could reasonably worry that the set of weak*-continuous linear functionals on $\mathcal{S}$ is larger than the set of weak*-continuous linear functionals on the larger set $\mathrm{Lip}(\mathbb{U}_0^1)$. However, the next result shows that $\mathcal{S}$ is weak* dense in bounded subsets of $\mathrm{Lip}(\mathbb{U}_0^1)$. As continuous functions are determined by their values on dense sets, the worry is not well-founded.

Lemma 7. *$\mathcal{S}$ is weak*-dense in norm bounded subsets of* $\mathrm{Lip}(\mathbb{U}_0^1)$.

Proof. From [28, Thm. 9], we know that, restricted to any compact subset of $\mathbb{U}_0^1$, $\mathcal{S}$ satisfies the conditions for the Stone-Weierstrass theorem. The compactness of finite sets and the equivalence of (b) and (c) in Lemma 6 shows that $\mathcal{S}$ is weak*-dense in norm bounded subsets of $\mathrm{Lip}(\mathbb{U}_0^1)$. $\square$

Institut für Mathematische Wirtschaftsforschung, Bielefeld University, Postfach 100131, 33501 Bielefeld, e-mail: mdumav@gmail.com, and Department of Economics, University of Texas, Austin, TX 78712-0301 USA, e-mail: max.stinchcombe@gmail.com