Ambiguity without a State Space

JOB MARKET PAPER

David S. Ahn*

November 2003

Abstract

Many decisions involve both imprecise probabilities and intractable states of the world. Objective expected utility assumes unambiguous probabilities; subjective expected utility assumes a completely specified state space. This paper analyzes a third domain of preference: sets of consequential lotteries. Using this domain, we develop a theory of Knightian ambiguity without explicitly invoking any state space. We characterize a representation that integrates a monotone transformation of first order expected utility with respect to a second order measure. The concavity of the transformation and the weighting of the measure capture ambiguity aversion. We propose a definition for comparative ambiguity aversion and uniquely characterize absolute ambiguity neutrality. Finally, we discuss applications of the theory: reinsurance, games, and a mean–variance–ambiguity portfolio frontier.

1 Introduction

Consider a terminally ill patient whose doctor suggests two treatments. The first is an established pharmaceutical. Numerous published studies concur that this drug is successful in thirty percent of cases. The second is a new experimental surgery. Preliminary trials for this surgery suggest a success rate between twenty and forty-five percent. The two treatments are mutually exclusive, so the patient must choose between them. Can we help the patient by framing her problem with either the von Neumann–Morgenstern (1944) or Savage (1954) theory of choice under uncertainty?

The former examines preferences for different probabilities over consequences, in this case, life and death. This assumes each choice induces a single objective probability on the outcomes. We cannot frame the patient's problem in the standard von Neumann–Morgenstern (henceforth vNM) setting: the probabilities associated with the surgery are ambiguous, since the patient is given a range of possible success rates. Other shortcomings of vNM theory, such as the Allais (1953)

*Stanford Graduate School of Business; ahn_david@gsb.stanford.edu; http://www.stanford.edu/∼ahnd. The author thanks Susan Athey, David Baron, Simon Board, Chris Chambers, Larry Epstein, Haluk Ergin, Yossi Feinberg, Peter Hammond, David Kreps, Mark Machina, Wojciech Olszewski, Andy Skrzypacz, and Chris Tyson for helpful comments and acknowledges a special debt to Bob Wilson for encouragement and support. Any further suggestions are welcome.

paradox, can be addressed by relaxing the independence axiom and using a more general utility function. These shortcomings fault the linear form of the expected utility function, not its domain. Our problem is more fundamental and exposes a deficiency in the primitives: the tacit assumption of precise probability. This deficiency is inherent to the domain of choice, so the theory cannot be salvaged by relaxing axioms.

Savage (1954) invented a subjective theory that lives in a richer domain and does not assume the probabilities of different outcomes are known.[1] Given the relevant states of the world, the decision maker conjectures a personal probability over the states and proceeds as if this conjecture is objective. Preference is defined over acts, which are assignments of consequences to states. To invoke the Savage machinery, our patient must be able to:

1. determine the relevant states of nature; and

2. decide how each choice assigns consequences to these states.

She fails both counts and cannot frame her problem using subjective utility theory. Regarding the state space, the patient has no medical training and no idea what the relevant states are. She knows only the information presented by her physician, expressed entirely in the space of probabilities over consequences. Even if she could comprehend the state space, the designers of the experimental surgery are unsure which states would make the surgery more likely to be successful. Studies are necessary exactly because no one knows which mapping from states to outcomes actually represents this new medical procedure. As with vNM theory, the problem with subjective utility theory is not any particular utility representation, but is fundamental to the structure of an act.

The machinery of vNM is too simple to express the patient's problem; the machinery of Savage is too complicated. This paper provides a third framework to analyze decisions under uncertainty. Unlike most research in decision theory, it does not propose an alternative utility function for either the vNM or Savage setting. Instead, it studies new primitives on the domain of choice: sets of lotteries over consequences. This domain incorporates ambiguity without appealing to a state space, thus avoiding the technology of states and acts. We can then express the patient's decision problem in formal terms: the established drug is represented as the singleton lottery that yields success with probability 0.3 and the experimental surgery as the set of lotteries that yield success with probabilities between 0.2 and 0.45.

Although our departure is less ambitious, its motivating spirit is similar to that of Gilboa and Schmeidler's theory of case–based decisions:

> Expected utility theory enjoys the status of an almost unrivaled paradigm for decision-making in the face of uncertainty. . . . While evidence has been accumulating that the theory is too restrictive (at least from a descriptive viewpoint), its various generalizations only attest to the strength and appeal of the expected utility paradigm. With few exceptions, all suggested alternatives retain the framework of the model, relaxing some of the more "demanding" axioms

[1] An important special case is the model of Anscombe and Aumann (1963), which we subsume in our discussion of general Savage theory.

> while adhering to the more "basic" ones. . . . Yet it seems that in many situations of choice under uncertainty, the very language of expected utility models is inappropriate. For instance, in many decision problems under uncertainty, states of the world are neither naturally given, nor can they be simply formulated (Gilboa and Schmeidler 1995, pp. 605–606).

The "basic" assumption we drop is the existence of a state space to capture ambiguity. Instead, we add sets of lotteries to the language of vNM theory to provide a new formulation of choice under ambiguity.

Ellsberg (1961) provided the classic illustration of ambiguity aversion. An urn contains ninety red, yellow, and black balls. There are thirty red balls in the urn. There are sixty yellow and black balls, but their specific proportions are unknown. A subject can bet on different colors, receiving \$100 if the chosen color is drawn from the urn. Ellsberg interviewed decision theorists and found most prefer a bet on red to a bet on yellow, yet also prefer a bet on yellow and black to a bet on red and black.

Savage's decision maker frames the problem with a state space of colors, $\{R, Y, B\}$, and a consequence space of monetary payoffs, $\{\$100, \$0\}$. She conceptualizes the bets as functions from the state space to the consequence space: betting on red maps $R \mapsto \$100$, $Y \mapsto \$0$, and $B \mapsto \$0$; betting on red and black takes $R \mapsto \$100$, $Y \mapsto \$0$, and $B \mapsto \$100$. When deciding between acts, she assigns probabilities to the three states space and utilities to the consequences, then integrates the utilities with respect to the assigned probabilities. The announced preference of red over yellow reveals a belief that a red ball is more likely than a yellow one. On the other hand, the preference of yellow and black over red and black indicates the opposite.

Although this procedure is normatively appealing, an actual decision maker might analyze the possibilities more simply, especially for decisions more complicated than the Ellsberg urn. As discussed above, the agent must perform two steps in implementing Savage's theory. First, she must conceptualize each state of the world, described as "a description of the world, leaving no relevant aspect undescribed" (1954, p. 9). In cases where the relevant aspects are complicated or unwieldy, describing all the states might be very difficult. Yet it is precisely these cases where the decision problem is rife with ambiguity. Second, she must decide how a prospect assigns consequences to these states of the world. The agent may be unsure of how a particular act assigns consequences to states, especially without empirical data on the realizations of past acts. Again, it is precisely such decisions that present the agent with ambiguity.

Therefore, framing the problem in the language of subjective expected utility immediately precludes two fundamental causes of ambiguity. Nonetheless, all existing theories of ambiguity live in some sort of subjective setting. They relax axioms to find a generalized utility representation to explain ambiguity aversion, but keep the fundamental domain of choice intact. For example, the maxmin model takes the minimum expected utility of an act over a set of priors on the state space (Gilboa and Schmeidler 1989), while the Choquet model takes the expected utility of an act with respect to a nonadditive probability capacity (Schmeidler 1989). Both assume the agent properly conceptualizes the state space and the mapping of consequences by acts. The only mechanism for ambiguity in these models is an improper assignment of probabilities to the states; otherwise the protocol for decision making remains similar to that of Savage. Yet the problem may not be the difficulty of assigning probabilities to states, but a more basic inability to conceptualize the state space in the first place.

We consider situations where the agent is unable to think of the decision problem in subjective terms. If she cannot implement the two mentioned steps, the primitives of Savage's model are not transparent. To borrow from Gilboa and Schmeidler, the "language" is fundamentally problematic. Unable to understand the states underlying her choices, the decision maker may still understand how her choices affect consequences, which are the ultimate objects of her utility. She can understand how something will make her feel, without understanding the causal mechanism or act that delivers that feeling. Instead of thinking of acts and states, the agent forms some boundaries on the possible consequential probabilities associated with each choice. For example, bets on the Ellsberg urn are compared by their corresponding ranges of possible odds. When she thinks of a bet on red, she does not conceptualize the function described above. Instead, she pictures the single lottery that yields \$100 with probability $\frac{1}{3}$ and \$0 with probability $\frac{2}{3}$. When she thinks of a bet on yellow, she pictures the entire set of lotteries that place between 0 and $\frac{2}{3}$ probability on winning \$100.

More generally, our decision maker has preferences over sets of lotteries. Each set captures the possible distributions on consequences associated with a particular option. Classic vNM theory is a special case where these sets are always singletons. This case attaches a single distribution to each choice, so ambiguity is never an issue. By enriching the domain of preference, we can introduce ambiguity in an objective setting.

Taking these sets as exogenous, our tack departs from the standard subjective approach, where there is no notion of objective or exogenous ambiguity. In a pure subjective theory, insofar as ambiguity exists, it is meaningful only in the mind of the agent. This austere view makes no additional assumptions of the world outside the agent's mind. While such parsimony is theoretically elegant, we believe there are compelling reasons to allow objective ambiguity.

Introducing objective ambiguity arms us with more detail, and this detail can capture realistic features of the decision problem. Fully subjective theories provide no device for the agent to incorporate outside information, or lack thereof, about uncertainty into her decision making. To illustrate, the ambiguity of the Ellsberg urn is objectively given by the experimental design. The subjects know the probability of a red ball is bounded below by zero and above by two thirds. At the same time, the subjects are also aware the number of red balls is objectively unambiguous, set by the experimenter as one third of the urn. We prohibit the agent from holding beliefs that contradict the information given by the experiment.

Also problematic is the inability to distinguish situations without any ambiguity and situations where ambiguity is resolved rationally by the agent. For example, the maxmin model represents attitude towards ambiguity by a set of multiple priors. The same set of priors also represents the existence of ambiguity. Then the presence of ambiguity is completely confounded with the agent's resolution of that ambiguity. If the agent is unsophisticated, we cannot determine if it is ambiguity or some other factor causing her irrational assessment. If she is sophisticated, it impossible to determine whether she resolved ambiguity rationally or there was never any ambiguity to be resolved. These factors can be separated only in a model where the the existence of ambiguity is exogenous and independent of its resolution. By starting with sets of lotteries, our model forces this separation.

Finally, we believe that avoiding a state space is not only substantively accurate, but also methodologically pragmatic. Our theory provides a formalism to incorporate ambiguity in applied models without modeling the state space explicitly. Almost all applied models are objective in probabilities. In fact, usually there is no state space at all. For example, one of the achievements of Harsanyi's model of games with incomplete information is that it transfers uncertainty over a state space to uncertainty over distributions of payoffs. We hope our model allows for applied analysis that incorporates ambiguity aversion, without the more complicated technology of subjective utility.

We close the introduction by discussing some related research. A form of our representation for conditional expectation was already proven and applied to Jeffrey's (1965) syntactic theory of decision by Bolker (1966, 1967), who deserves original credit for the mathematical result.[2] We first comment on some technical differences. Bolker assumes a nonatomic complete Boolean algebra; we allow atoms, but require a divisibility axiom on regular sets. We also provide more structure on the final objects: the Radon–Nikodym derivative is continuous and both measures have full support. Within this narrower class, our conditions become necessary, as well as sufficient, for the representation. Finally, while they share a critical geometric insight, the proofs are significantly different, since we invoke recent results on λ–systems.

Substantively, our theory differs from Jeffrey's theory of joint desirability and probability over logical propositions or sentences, which is not a treatment of ambiguity. His framework does not immediately translate to familiar economic formulations, since a proposition can be interpreted as a consequence, as a state, and as an act.[3] In terms of mechanics, the probability on propositions is the only carrier of uncertainty and is not allowed to vary. We consider the space of *all* possible lotteries on consequences, but identify a fixed measure over the various lotteries. Perhaps our theory can be viewed as a form of Jeffrey's that considers "propositions" regarding risk: both theories have preferences for information, captured as a sentence about the world or as a set of lotteries, and implement some form of conditional expectation. That said, we suspect Jeffrey would object to our basic model: "I take it to be the principal virtue of the present theory, that it makes no use of the notion of a gamble or of any other causal notion (Jeffrey 1965, p. 147)."

Preferences over sets of lotteries appear in theories of flexibility (Dekel, Lipman, and Rustichini 2001) and self-control (Gul and Pesendorfer 2001). However, their interpretation of sets is fundamentally different. They view sets as menus of stochastic choices; at an implicit second stage, the agent chooses a single lottery from the menu. For us, a set reflects objective information about the risks involved in a decision—there is no second stage of choice.

---

[2]We are indebted to Larry Epstein for bringing Bolker's work to our attention, and thank Chris Chambers for subsequent references on the Jeffrey model.

[3]Nonetheless, there is at least one application of the theory to welfare economics (Broome 1990).

Two other papers use sets of lotteries to capture ambiguity, but with different utility representations. Olszewski (2003) characterizes a generalized form of α–maxmin utility, which evaluates a set by a convex combination of its minimal and maximal elements and is further discussed in Section 3. Stinchcombe (2003) presents a dual form of the Expected Utility Theorem on the mixture space of sets. An advantage of these linear approaches is that the resulting utilities are well-defined over lower dimensional sets, which our measure-theoretic approach is forced to ignore. On the other hand, their linear methods are finite-dimensional and assume finite consequences, while our arguments apply to continuous consequences spaces. We should also note that neither paper shares our philosophical position against the descriptive veracity of the Savage formulation in the presence of ambiguity, and should not be held responsible for our motivations or opinions.

Another recent strand of research enriches the subjective model with exogenous sets of *priors* on the state space (Damiano 1999, Gajdos, Tallon, and Vergnaud 2003, Hayashi 2003, Wang 2003). These papers retain the basic infrastructure of states and acts, which we oppose. A possible technical reconciliation is to assume a generalized form of probabilistic sophistication, where an ambiguous act is evaluated by its induced set of distributions over consequences; this is further discussed when the model is formally introduced in Section 2. Both approaches take the set of lotteries or priors as exogenous; Siniscalchi (2003) proposes an axiomatic identification of plausible priors in a subjective model using only behavioral information.

Ergin and Gul (2002), Klibanoff, Marinacci, and Mukerji (2003), and Nau (2003) characterize transformed utility functions similar to Corollary 4, but in subjective models with product state spaces. Taking each dimension as a stage of uncertainty, preferences are defined over second order acts. Our domain is sets of *simple* lotteries, not lotteries over lotteries; there is no explicit or verifiable second order uncertainty in the model. For example, we do not allow bets on the measure over lotteries, and cannot elicit information about ambiguity by isolating and varying outcomes on second order uncertainty. We finally reiterate the most important distinction: our model is designed specifically to avoid the use of state and acts altogether.

In the next section, we formally introduce the primitives of our theory. Section 3 characterizes maxmin utility in our setting. Section 4 contains our main representation: the decision maker integrates a transformed expected utility with respect to a second order measure, conditioning on the objective set of lotteries. Section 5 discusses comparative and absolute ambiguity aversion and conducts some comparative statics. Section 6 presents three applications: reinsurance, games, and a mean–variance–ambiguity frontier.

2 Model

The compact Polish space X denotes the set of deterministic outcomes.[4] ΔX is the set of Borel probability measures on X, endowed with the Prohorov metric ρ, which induces the topology of weak convergence. $\Delta^2 X = \Delta(\Delta X)$ is the set of Borel probability measures on ΔX, or second

[4]A metric space is Polish if it is complete and separable.

order measures on X. We refer to elements of ΔX as "lotteries," even though they may not have finite support, and reserve the term "measures" for elements of $\Delta^2 X$. $\mathcal{K}(\Delta X)$ denotes the family of nonempty closed subsets of ΔX. This family is endowed with the Hausdorff metric:

$$d(A,B) = \max\left\{\max_{a\in A}\min_{b\in B}\rho(a,b)\,,\,\max_{b\in B}\min_{a\in A}\rho(a,b)\right\}.^5$$

When X is finite with n elements, the topology of weak convergence on ΔX is homeomorphic to the Euclidean topology on $\{x \in \mathbb{R}^{n-1} : x_1,\ldots,x_{n-1} \geq 0, \sum_{i=1}^{n-1} x_i \leq 1\}$.

We also consider the space of nonempty regular and singleton subsets of ΔX, denoted

$$\mathcal{K}^*(\Delta X) = \{A \subseteq \Delta X : \overline{\mathrm{int}(A)} = A \text{ or } |A| = 1\} \setminus \emptyset.^6$$

This refinement to regular sets is explained in Section 4. In brief, we wish to exclude pathologies of measure zero. At the same time, the relevant measure is identified by the representation; so we cannot assume an exogenous family of null sets. Regularity provides a topological solution to this measure theoretic problem, which we believe is a novel contribution of the model.

Finally, $\succsim$ is a complete and transitive binary relation on either $\mathcal{K}(\Delta X)$ or $\mathcal{K}^*(\Delta X)$, with $\succ$ and $\sim$ having the standard definitions.

The set of lotteries represents the possible risks associated with a particular choice. An ambiguous set captures the imprecision of the probabilities associated with a particular choice. For example, in the Ellsberg urn, the agent knows that a bet on yellow is associated with a range of lotteries, represented as the interval $[0, \frac{2}{3}]$, but has no further information beyond that range. On the other hand, a bet on red is associated with the single lottery $\{\frac{1}{3}\}$. Any singleton is unambiguous, because the risk is known and precise.

Definition 1. A set A is *unambiguous* if $|A| = 1$.

We view this simple definition as a strength of the model.[7]

---

[5] d is a well defined metric over the family of closed sets.

[6] A set is *regular* if it is equal to the closure of its interior. This should not be confused with the separation condition for a regular topological space.

[7] Ghirardato and Marinacci (2002) provide a simple definition of unambiguous acts as constant acts, but beg what constitutes an unambiguous event. Epstein and Zhang (2001, p. 273) define an event T in the state space as unambiguous if:

1. For all disjoint subevents A, B of $T^{\complement}$, acts h, and outcomes $x^*, x, z, z' \in X$,

$$\begin{pmatrix} x^* & \text{if } s \in A \\ x & \text{if } s \in B \\ h(s) & \text{if } s \in T^{\complement} \setminus (A \cup B) \\ z & \text{if } s \in T \end{pmatrix} \succsim \begin{pmatrix} x & \text{if } s \in A \\ x^* & \text{if } s \in B \\ h(s) & \text{if } s \in T^{\complement} \setminus (A \cup B) \\ z & \text{if } s \in T \end{pmatrix}$$

implies

$$\begin{pmatrix} x^* & \text{if } s \in A \\ x & \text{if } s \in B \\ h(s) & \text{if } s \in T^{\complement} \setminus (A \cup B) \\ z' & \text{if } s \in T \end{pmatrix} \succsim \begin{pmatrix} x & \text{if } s \in A \\ x^* & \text{if } s \in B \\ h(s) & \text{if } s \in T^{\complement} \setminus (A \cup B) \\ z' & \text{if } s \in T \end{pmatrix};$$

We can think of this objective setting as the extension of the distributional approach in the presence of ambiguous information. Suppose the agent is probabilistically sophisticated, in the sense introduced by Machina and Schmeidler (1992). A probability assessment μ over states is naturally mapped by an act f to a probability assessment ν over consequences by its distribution: $\nu(S) = \mu(f^{-1}(S))$. This distribution is sufficient to characterize her preference, since her utility for an act is determined exactly by its distribution over outcomes. If two acts induce the same distribution on outcomes, then the agent will be indifferent between them. An act contains more information than is really needed in a theory of choice; all payoff relevant information is captured by its distribution. We believe that it is possible to develop a parallel theory of choice under ambiguity without appealing to a state space. Ours is a theory of ambiguity where the only relevant information is captured by the *set* of possible distributions. A set Π of probability assessments over the states induces a set A of distributions or laws over the consequences: $A = \{\nu : \nu(S) = \mu(f^{-1}(S))$ for some $\mu \in \Pi\}$. If two ambiguous acts induce the same sets of distributions, then they are equally appealing.

This extends the notion of distributional evaluation from *single* distributions induced by unambiguous acts to *sets* of distributions induced by ambiguous acts. With or without ambiguity, if two acts yield the same sets of possible distributions, they should be considered equally desirable. For example, in the Ellsberg urn, our model implicitly assumes the agent treats yellow and black symmetrically and is indifferent between them. Absent further information, we find this imposed symmetry a reasonable assumption. Ellsberg himself reported, "In our examples, actual subjects do tend to be indifferent between betting on [yellow or black]. . . . the reasons, if any, to favor one or the other balanced out subjectively so that the possibilities entered into their final decisions weighted equivalently (Ellsberg 1961, p. 658)."

Thinking in terms of distributions illuminates the connection between the traditional subjective approaches and our objective model. The Savage model immediately translates to the objective von Neumann–Morgenstern setting once the personal probability is fixed, because each act maps to a singleton lottery over the consequences. Our model can interpreted as an objective version of a multiple priors model. These sets of lotteries are like those generated by an exogenous set of multiple priors.

This model sharply delineates ambiguity as a closed set of lotteries. In reality, the decision maker may not have such crisp boundaries on the possible lotteries. Instead, she may think a variety of sets are possible, and have a belief on the likelihood of these sets, an element of $\Delta(\mathcal{K}(\Delta X))$. Moreover, if the agent can hold ambiguous beliefs about the consequences ΔX, then she may also hold ambiguous beliefs about the ambiguity, captured as $\mathcal{K}(\Delta(\mathcal{K}(\Delta X)))$. Iterative applications of risk $\Delta(\cdot)$ and ambiguity $\mathcal{K}(\cdot)$ produce infinite levels of ambiguity about ambiguity. If these levels of higher order ambiguity collapse to a single expanded space of consequences, then our model loses no generality in considering only "first order" ambiguity. We prove the hypothesis in (Ahn 2003),

2. The condition obtained if T is everywhere replaced by $T^{\complement}$ in the prior statement is also satisfied.

which constructs a universal type space of ambiguous beliefs that can also serve as a universal consequence space for this model.

3 Maxmin representation

Given an objective or exogenous set of probabilities on the state space, maxmin expected utility is one particular way of comparing acts. By no means it is the only way to make these comparisons, or even the most plausible. Still, we begin by providing an axiomatic derivation of maxmin expected utility in our objective setting. We present the theorem as a straw person; we are not arguing the descriptive power of these axioms. On the contrary, we find one of the axioms particularly distressing. Later, these preferences will serve as a useful benchmark for our alternative representation.

The theorem of this section translates the seminal result of Gilboa and Schmeidler (1989, Theorem 1) to our setting. They provide axioms in an Anscombe–Aumann setting that characterize the following utility for an act $f : S \to \Delta X$:

$$U(f) = \min_{P \in C} \int u \circ f \, dP,$$

where u is linear in probabilities, and C is a closed convex set of measures on the state space S.[8]

Axiom 1 (Continuity). $\{B : B \succsim A\}$ and $\{B : A \succsim B\}$ are closed (in the Hausdorff metric topology).

This definition is standard.

Axiom 2 (Decreasingness). $A \subseteq B$ implies $A \succsim B$.

Axiom 2 captures an extreme form of ambiguity aversion. The agent is always worse off when more lotteries are added to a set A, even if these lotteries are individually preferred to A. For example, suppose the agent has just bet on the yellow ball from the Ellsberg urn. Then knowing there are between zero and sixty yellow balls can be no better than knowing there are no yellow balls at all. This condition seems unreasonable as a primitive assumption on the agent's preference, but that is exactly our point. It is a necessary consequence of a maxmin criterion, sharply demonstrating an unrealistic feature of that theory in our objective setting.

Axiom 3 (Disjoint upper closure). $A \succsim C$, $B \succsim C$, and $A \cap B = \emptyset$ imply $A \cup B \succsim C$.

Mathematically, Axiom 3 means the upper contour sets of $\succsim$ are closed under disjoint union. The dual property, closure of the lower contour sets under disjoint union, is implied by Axiom 2. Economically, upper closure tempers the extreme ambiguity aversion captured by Axiom 2. Adding

[8] This representation is characterized for general Savage acts by Casadesus-Masanell, Klibanoff, and Ozdenoren (2000) and in a subjective setting with exogenous sets of priors on the state space and an "anchor" prior for each set by Wang (2003).

more lotteries to a set is not detrimental if the additional lotteries would individually be preferred to the original set.

These axioms are necessary and sufficient for maxmin utility in our setting.

Theorem 1. *$\succsim$ on $\mathcal{K}$ meets Axioms 1, 2, and 3 if and only if there exists a continuous function $u : \Delta X \to \mathbb{R}$ such that*

$$U(A) = \min_{a \in A} u(a)$$

is a utility representation of $\succsim$.

Proof. See Appendix A.1.

Unlike the result of Gilboa and Schmeidler (1989, Theorem 1), this result does not impose any sort of independence axiom. While the function u on lotteries is continuous, it may not have linear indifference curves. Therefore, our result allows the agent to be unsophisticated in both her treatment of probability and her treatment of utility, in the sense that she violates the independence axiom. So, she can be subject to the Allais paradox and the Ellsberg paradox. Of course, additional structure on the preferences delivers linearity.

Axiom 4 (Singleton independence). $\{a\} \succsim \{b\}$ if and only if $\{\alpha a + (1-\alpha)c\} \succsim \{\alpha b + (1-\alpha)c\}$ for all $c \in \Delta X$, $\alpha \in (0,1)$.

This is the standard independence axiom, restricted to the singletons given our generalized domain. In our context, it is analogous to the Certainty–Independence axiom of Gilboa and Schmeidler (1989, Axiom A.2).

Corollary 2. *$\succsim$ meets Axioms 1–4 if and only if there exists a linear $u : \Delta X \to \mathbb{R}$ such that*

$$U(A) = \min_{a \in A} u(a)$$

is a utility representation of $\succsim$.

The clean functional form and crisp axiomatic characterizations make maxmin utility appealing. Nonetheless, we find this form of utility disturbing, especially the necessity of Axiom 2. Aside from the minimal lottery, the representation ignores all the other lotteries included in a set A. Indeed, Ellsberg anticipated this result with dissatisfaction: "In almost no cases . . . will the *only* fact worth noting about a prospective action be its 'security level': the 'worst' of the expectations associated with reasonably possible probability distributions. To choose on a 'maxmin' criterion alone would be to ignore entirely those probability judgments for which there is evidence (Ellsberg 1961, p. 662)."

In part to mitigate this extreme form of ambiguity aversion, a variant of maxmin utility takes a weighted combination of the worst and best distributions in a set. This is sometimes called α–maxmin utility and is axiomatized by Ghirardato, Maccheroni, and Marinacci (2002) in a subjective setting and by Olszewski (2003) in an objective setting similar to ours. While α–maxmin utility improves simple maxmin, it retains some problems. For example, α–maxmin utility still ignores almost all of the information contained in the set of priors or the set of lotteries; preferences are completely characterized by minimal and maximal elements. Consider the lotteries over \$0 and \$100, represented on $[0,1]$ by their probabilities for \$100. Then α–maxmin utility is indifferent between $[0,0.5]\cup[0.9,1]$ and $[0,0.1]\cup[0.5,1]$, while the latter seems more intuitively appealing. We desire a utility that incorporates information about all the priors or lotteries in a set.

4 Main representation

Given the extreme nature of the representation in Theorem 1, a more moderate resolution of ambiguity is desirable. Consider a utility function u on the single lotteries and a probability measure μ on the Borel subsets of ΔX. We believe the following utility representation is a more reasonable resolution of ambiguity:

$$U(A) = \frac{\int_A u\, d\mu}{\mu(A)}.$$

The agent conditions her utility on a second order measure, given the information that the set A of lotteries obtains. This incorporates every lottery in A, weighted by the measure μ.

Additionally, the utility function should be continuous and the measure should be nonatomic. Continuity of the preference would provide both conditions. Unfortunately, the proposed utility violates continuity when defined over all closed sets of lotteries. Suppose the lotteries are over two outcomes, represented by $[0,1]$. Set $u(x)=x$ and μ to the Lebesgue measure. Then the sets $A_\delta = [\frac{1}{4}-\delta, \frac{1}{4}+\delta]\cup[\frac{3}{4}-2\delta, \frac{3}{4}+2\delta]$ and $B_\delta = [\frac{1}{4}-2\delta, \frac{1}{4}+2\delta]\cup[\frac{3}{4}-\delta, \frac{3}{4}+\delta]$ both converge to the doubleton $\{\frac{1}{4}, \frac{3}{4}\}$ as $\delta \to 0$. Yet $U(A_\delta) = \frac{7}{12}$ and $U(B_\delta) = \frac{5}{12}$ for all δ, so continuity fails. Part of the problem is that $\{\frac{1}{4}, \frac{3}{4}\}$ is a null set, so its conditional expectation is undefined.

This motivates our construction of $\mathcal{K}^*$, which limits attention to regular sets and singletons. $\mathcal{K}^*$ excludes pathological sets like $\{\frac{1}{4}, \frac{3}{4}\}$ on topological criteria, without fixing μ a priori. Of course, we include the singletons to retain unambiguous choices. The decision maker must have either ambiguous marginals across all dimension or no ambiguity at all; she cannot know precisely the probabilities of some outcomes but not of others.[9] Throughout this section, we will restrict $\succsim$ to be a binary relation on $\mathcal{K}^*$, rather than $\mathcal{K}$.[10]

Our representation involves measures, so we need closure under standard set theoretic operations. Regular sets are not closed under these operations: consider the intersection of two regular sets who meet only at their boundaries. Regular sets are closed under the regularized set operations,

[9] Another way of thinking about the restriction, pointed out to us by Wojciech Olszewski, is: if X is finite, then regular sets are of the same dimension as ΔX.

[10] Rather than refining the domain of choice, a different approach might invoke a lexicographic probability system (Blume, Brandenburger, and Dekel 1991) of measures.

defined as:

$$\begin{aligned} A \cup' B &= \overline{\text{int}(A \cup B)}; \\ A \cap' B &= \overline{\text{int}(A \cap B)}; \\ A \setminus' B &= \overline{\text{int}(A \setminus B)}. \end{aligned}$$

We slightly abuse notation and drop the primes: all set operations in the following definitions are regularized. We can now introduce the axioms.

Axiom 5 (Continuity*). $\{B \in \mathcal{K}^* : B \succsim A\}$ and $\{B \in \mathcal{K}^* : A \succsim B\}$ are closed (in the relative Hausdorff metric topology).

Axiom 6 (Disjoint set betweenness). Suppose $A \cap B = \emptyset$. $A \succsim (\succ) B$ implies $A \succsim (\succ) A \cup B \succsim (\succ) B$.

Gul and Pesendorfer (2001) assume a similar axiom in their work on self control. Our axiom is technically weaker in one sense, applying only to disjoint unions; it is stronger in another, preserving both weak and strict preference. More importantly, set betweenness has a different substantive interpretation in our model. Gul and Pesendorfer think of set betweenness in the context of temptation: unchosen or suboptimal elements of a menu $A \cup B$ carry a disutility of temptation. Our sets are not menus, but provide information about possible lotteries. Consider the following phrases: I have some good news (A); I have some bad news (B); and I have some good news and some bad news ($A \cup B$). Hearing only good news is strictly better than hearing good news with bad news, which is itself strictly better than hearing only bad news.

The maxmin utility characterized in Theorem 1 barely fails this condition. Instead, maxmin utility implies the following: if $A \succ B$, then $A \succ A \cup B \sim B$. An agent with maxmin utility ignores all the good news (A), paying attention only to the bad news (B). Disjoint set betweenness distinguishes our theory by forcing the agent to pay attention to good news. Later, we show that maxmin utility is a limit case of our representation in a precise topological sense.

Axiom 7 (Balancedness). Suppose $[A \cup B] \cap [C \cup D] = \emptyset$ and $A \sim B \succ C, D$ (or $A \sim B \prec C, D$). $A \cup C \succsim B \cup C$ implies $A \cup D \succsim B \cup D$.

Balancedness has the flavor of the sure-thing principle (Savage 1954, Section 2.7). Consider the sets $A \cup C$ and $B \cup C$ in the hypothesis. These two unions share C and differ only in that one has A and the other has B. Then, if we take any other $D \prec A$, this preference is preserved. The only data that matter in evaluating these types of sets are the places where they differ, namely A and B. Their intersection, D in the conclusion, does not affect their relative desirability. Our axiom is even weaker because we add the additional restriction that $A \sim B$. Although loosely related, balancedness is *not* the same as the sure-thing principle. The indifference and disjointedness assumptions really have no analog in the Savage setting. Conversely, the sure-thing principle has no direct translation in our setting.

Axiom 8 (Divisibility). For any regular A, there exists $A' \subsetneq A$ such that $A' \sim A$ and $A' \cup B \sim (A \setminus A') \cup B$ for some $B \nsim A$ with $A \cap B = \emptyset$.

Divisibility is a technical axiom which allows us to identify the underlying measure μ. It splits ΔX into smaller and smaller partitions of indifferent sets. Similar assumptions are ubiquitous in the axiomatic treatment of probability. For example, Savage assumes the state space is fine and tight, forcing it to be uncountably rich. Qualitative order axioms are generally insufficient to identify a quantitative probability on a finite algebra (Kraft, Pratt, and Seidenberg 1959).

These axioms are necessary and sufficient for the proposed utility representation.

Theorem 3. *$\succsim$ on $\mathcal{K}^*$ meets Axioms 5–8 if and only if there exists a continuous $u : \Delta X \to \mathbb{R}$ and a nonatomic probability measure μ on ΔX with full support such that*

$$U(A) = \frac{\int_A u \, d\mu}{\mu(A)}$$

is a utility representation of $\succsim$.[11]

Proof. See Appendix A.3.

Remark. Making no restrictive assumptions on the shape of risk, the nonparametric domain of choice presented here considers arbitrary distributions on X. However, the proof only assumes that ΔX is a compact Polish mixture space. The characterization remains valid when restricted to a family of parameterized distributions, provided the parameter space is a compact Polish mixture space. This is true even if the consequence space X is not compact or Polish. For example, if the decision maker knows that the risk is normal over all levels of wealth, we can define preference over closed subsets of a compact rectangle $M \times V \subset \mathbb{R}^2$, associated with normal distributions of different means M and variances V, while $X = \mathbb{R}$ is not compact.

On the singletons, this functional form is undefined, but the continuity axiom and Lemma 19 of Appendix A.3 imply $U(\{x\}) = u(x)$, which we adopt as convention. As with Theorem 1, we do not assume the agent's utility over singleton lotteries is linear. Singleton independence provides further structure on the utility function.

Corollary 4. *$\succsim$ on $\mathcal{K}^*$ meets Axioms 4–8 if and only if there exists a linear $u : \Delta X \to \mathbb{R}$, strictly increasing and continuous $\phi : \mathbb{R} \to \mathbb{R}$, and a nonatomic probability measure μ on ΔX with full support such that*

$$U(A) = \frac{\int_A \phi \circ u \, d\mu}{\mu(A)}$$

is a utility representation of $\succsim$.

Imposing singleton independence, we retain standard expected utility as a special case on the unambiguous singletons. The utility u is a classic linear expected utility function, and ϕ is a

[11] As mentioned in the introduction, a version of this result was already proven by Bolker (1966).

transformation that retains the ordinal independence condition on preferences; both u and $\phi \circ u$ produce linear indifference curves. The classic Expected Utility Theorem states there exists *some* linear utility representation, but not that *all* utility representations must be linear. There are nonlinear utility representations of independent preference, as shown in Figures 1 and 2. The classic theory is cardinal with respect to the value function on deterministic outcomes X (up to scale transformations), but is only ordinal with respect to the utility function on lotteries ΔX. The functional form of utility used to represent preferences over single lotteries is important when we integrate that form over sets of lotteries.

Although related, the curvature of utility in our model is not induced in the traditional sense of direct second order risk aversion on compound or two–stage lotteries. This would suggest a relaxation of the reduction axiom in the space of compound lotteries, about which we have nothing to say. The domain of our model is *sets* of simple lotteries, rather than *lotteries* of simple lotteries. The measure μ over lotteries is fixed by the representation; it is not assumed as a primitive of the theory, nor is it allowed to vary to reflect different second order uncertainties. Instead, the cardinality of $\phi \circ u$ is induced by the agent's preferences over subsets of ΔX.

Thus one part of an agent's attitude towards ambiguity is the transformation ϕ, which captures a cardinal intensity of preference one type of lottery compared to another. This intensity is important because the utility integrates this intensity with respect to the (endogenous) measure μ. The manifestation of ambiguity aversion as a departure from linearity should be comforting, since it resonates the traditional analysis of risk aversion. We develop this metaphor more later.

The second part of her attitude is determined by her weighting μ. One interpretation of our setting is a game against nature, where nature decides which lottery is actually realized.[12] Then μ is the agent's belief about nature's mixed strategy. If the agent thinks μ puts more probability on worse lotteries, she thinks nature will more likely choose a bad lottery. The measure μ can be interpreted as an agent's assessment of her "luck" in ambiguous situations. The more weight she places on better lotteries, the luckier she believes herself to be. Two agents can share a common utility function u and transformation ϕ, but still have different preferences because one thinks of herself as luckier than the other. The maxmin utility characterized in Theorem 1 is an extreme case where the agent pessimistically believes that nature always chooses the minimal element of a set.[13] Here, our decision maker weights all the possible lotteries in a set by μ, using all of the available information.

Another interpretation of μ is less wedded to the statistical view of the decision problem as a game against nature. We can interpret μ as a measure of salience or how much attention the agent pays to the various lotteries. Worse lotteries loom larger in the minds of those with a distaste for ambiguity. The measure μ then corresponds to the personal attention given to the possible lotteries, normalized so $\mu(\Delta X) = 1$.

Similar concepts of luck are missed in a subjective model where the agent takes a fixed weighted

[12] This interpretation is pushed more explicitly by Olszewski (2003).

[13] Of course, this extreme sense of bad luck cannot actually be defined as a probability measure over nature's behavior.

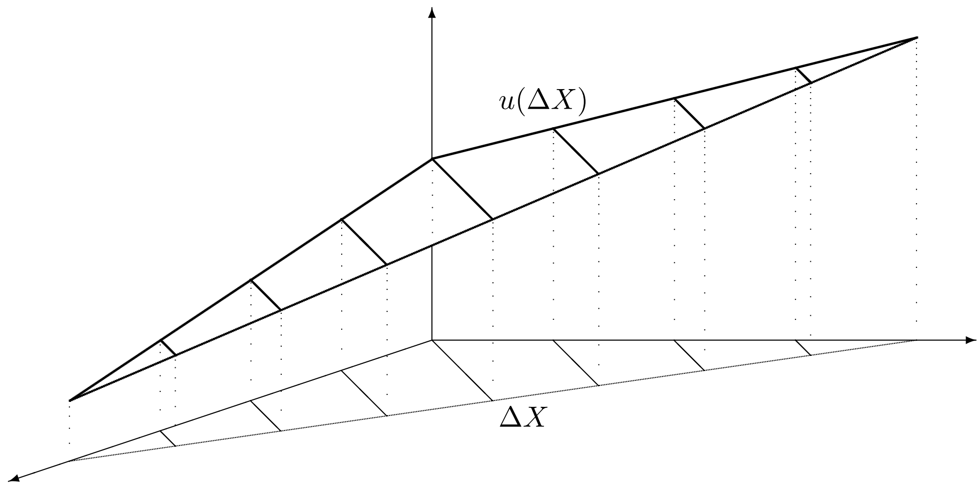

Figure 1: Linear expected utility (u)

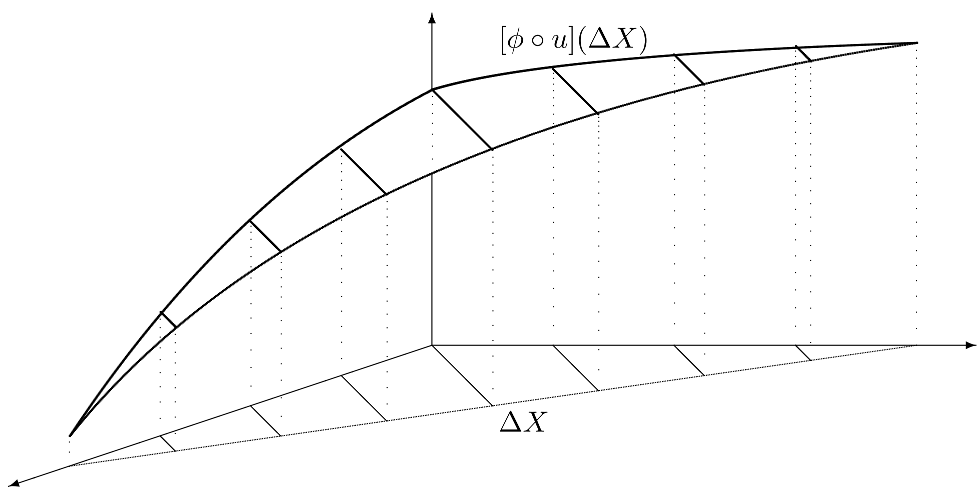

Figure 2: Nonlinear "expected" utility ($\phi \circ u$)

In Figure 1, the simplex of probability distributions ΔX lies on the ground, and the linear expected utility curve $u(\Delta X)$ floats above it. Since u represents independent preference, it has linear indifference curves. Figure 2 shows the same independent preference, except now the linear utility representation u is transformed by some concave function $\phi : \mathbb{R} \to \mathbb{R}$. Notice that the transformed utility retains the same linear indifference curves.

average over her priors. This is because whether the weighting is optimistic or pessimistic depends on the act being evaluated. For example, if she holds a bet on yellow in the Ellsberg urn, placing more weight on distributions with many yellow balls is optimistic. On the other hand, if she holds a bet on black, the same weighting would be considered pessimistic.

5 Ambiguity aversion

5.1 Comparative ambiguity aversion

From here on, we take the representation of Theorem 3 as given. In this section, we develop tools to discuss ambiguity aversion in our objective setting. We first introduce concepts to compare ambiguity aversion across individuals.

Definition 2. Suppose $\succsim_1$ and $\succsim_2$ are equivalent on ΔX. $\succsim_1$ is *locally more ambiguity averse at* A than $\succsim_2$ if

$$A \succsim_1 (\succ_1)\{a\} \Rightarrow A \succsim_2 (\succ_2)\{a\}.$$

$\succsim_1$ is *(globally) more ambiguity averse* than $\succsim_2$ if it is locally more ambiguity averse at all A.

This definition is the objective analog of Epstein's (1999) definition of comparative ambiguity aversion. He considers $\succsim_1$ more ambiguity averse that $\succsim_2$ if for every arbitrary act f and any unambiguous act g,

$$f \succsim_1 (\succ_1) g \Rightarrow f \succsim_2 (\succ_2) g,$$

where an act is considered unambiguous if it is measurable with respect to a λ-system of unambiguous events.[14] The definition by Ghirardato and Marinacci (2002) is identical, except g in the hypothesis is further restricted to be a constant function. The disagreement on what exactly constitutes an ambiguous act leads to these disparate definitions. Here, singleton lotteries take the place of λ-measurable or constant acts and sets of lotteries take the place of arbitrary acts, allowing us to finesse the issue.

Notice we begin by fixing preferences over the singleton lotteries. This is necessary for the same reasons that analysis of risk aversion must assume fixed preferences over sure consequences, usually a monotonic preference for money. If X is wealth, then this is equivalent to saying the agents are equally risk averse. But, our definition also applies if X is finite or multidimensional and there is no obvious notion of risk aversion. Our definition works even if agents do not meet the singleton independence condition, as long as they share the same preferences over the singletons.

Given that two agents share risk preferences, we can consider a_1 and a_2 their respective ambiguity-free equivalents to A if $\{a_1\} \sim_1 A$ and $\{a_2\} \sim_2 A$.[15] These ambiguity-free equivalents are conceptually similar to certainty equivalents in the theory of risk aversion. An agent is more risk averse than another if her certainty equivalent for a gamble is lower than another's. Here,

[14] See Definition 6 in Appendix A.3 for the definition of a λ-system.

[15] More generally, we can consider the entire set $\{a \in \Delta X : \{a\} \sim A\}$ as the ambiguity-free equivalent. This set is a hyperplane if $\succsim$ meets singleton independence.

we replace the natural ordering on money with the preference ordering on singletons. Then agent 1 is more ambiguity averse at A than agent 2 if her ambiguity free equivalent $x_1 \prec x_2$, agent 2's equivalent. An agent can be more ambiguity averse for some sets but less ambiguity averse for others, in the same way she can be more risk averse on some range of income and less risk averse on another.

We mentioned in Section 4 that maxmin utility nearly meets the axioms of Theorem 3. Specifically, it only barely fails to meet disjoint set-betweenness. We now show that maxmin utility is a limit case of our representation. To consider limits in the space of possible preferences, we define a topology on the space of preferences. For a fixed set $A \in \mathcal{Z}$, let

$$d_A(\succsim, \succsim') = \min\{\rho(x, x') : x \sim A, x' \sim' A\}$$

recalling ρ is the Prohorov metric.[16] We now say $\succsim_n \to \succsim$ if $d_A(\succsim_n, \succsim) \to 0$ for all $A \in \mathcal{Z}$. So, a sequence of preferences has a limit if the respective ambiguity-free equivalents for A approach a limit for any set $A \in \mathcal{Z}$.[17]

Remember our discussion of the two sides of ambiguity aversion: the transformation ϕ reflecting a cardinal utility towards gambles and the probability assessment μ reflecting an attitude about one's luck. To conduct comparative statics, we isolate each effect by keeping the other fixed. We begin by fixing the measure μ and comparing the curvature of ϕ. In the theory of risk, more curvature corresponds to more risk aversion. Similarly, in our theory more curvature corresponds to more ambiguity aversion. For a fixed risk profile, we let $\succsim_{\text{MMEU}}$ refer to the corresponding maxmin expected utility.

Proposition 5. *Suppose $\{\succsim_i\}$ have representations $\{(u, \phi_i, \mu)\}$. $\phi_1 = h \circ \phi_2$ for some concave and strictly increasing $h : \mathbb{R} \to \mathbb{R}$ if and only if $\succsim_1$ is more ambiguity averse than $\succsim_2$.*

Further, if ϕ_i is continuously differentiable with

$$\min_x -\frac{\phi_i''(x)}{\phi_i'(x)} \to \infty,$$

then $\succsim_i \to \succsim_{MMEU}$.[18]

Proof. The first part of the proposition follows almost directly from Jensen's Inequality. For the second part, make u positive by adding a sufficiently large constant. The result follows by taking a subsequence $\phi_{n(m)}$ with $\phi_{n(m)}$ a concave transformation of $-e^{-mx}$, then taking $m \to \infty$. □

The fraction $-\frac{\phi_i''(x)}{\phi_i'(x)}$ strongly resembles the standard Arrow–Pratt measure of absolute risk aversion. Fixing a probability μ, we can construct a similar quantitative measure to compare ambiguity aversion. As with risk, the ratio of second to first derivative provides a measure of curvature and captures the level of ambiguity aversion for the agent. As this measure approaches

[16] d_A is a semimetric on the the space of preferences.
[17] Notice that this convergence need not be uniform across A.
[18] Analogous results in a subjective setting are provided by Klibanoff, Marinacci, and Mukerji (2003)

infinity, the agent's preferences get closer to maxmin utility. Her relative distaste for worse lotteries increases, and she wishes more and more to avoid sets that include such lotteries.

Now we fix the transformation ϕ and vary the probability μ. Recall μ captures the agent's perception of her luck. The partial order of stochastic dominance formalizes what it means for one agent to consider herself "luckier" than another. A measure stochastically dominates another precisely if it puts more weight on higher or better lotteries. We let $\Delta A = \{\mu \in \Delta X : \mu(A) = 1\}$ for any set $A \subseteq X$. Let $\min_{\succsim} A = \{a \in A : \{a\} \precsim \{b\} \text{ for all } b \in A\}$ refer to the $\succsim$-minimal elements of A. $\min_{\succsim} A$ is closed, so $\Delta(\min_{\succsim} A)$ is also closed. Finally, $\mu|A$ is the conditional probability $[\mu|A](S) = \frac{\mu(A\cap S)}{\mu(A)}$.

Proposition 6. *Suppose $\{\succsim_i\}$ have representations $\{(u, \phi, \mu_i)\}$. $\mu_2|A$ stochastically dominates $\mu_1|A$ with respect to the lattice $\succsim_1 |A$ if and only if $\succsim_1$ is locally more ambiguity averse at A than $\succsim_2$.*

Further, if $\mu_i|A \to \Delta(\min_{\succsim} A)$ (in the topology of weak convergence on $\Delta^2 X$) for all A, then $\succsim_i \to \succsim_{MMEU}$.

Proof. This follows directly from definitions. □

This is another way maxmin utility is a limit case of our representation. The more unlucky an agent considers herself, the closer her behavior is to maxmin preference. She focuses more and more of her attention on the worse lotteries, until the worst lottery becomes the sole criterion for comparison.

5.2 Absolute ambiguity neutrality

Having established these relative definitions, we introduce an absolute benchmark for ambiguity neutrality. Our aim is to axiomatically identify a unique representation for ambiguity neutrality. In the subjective literature, there is a natural benchmark of probabilistic sophistication, which we do not have in our setting. In the theory of risk aversion, risk neutrality is identified by a linearity in the Bernoulli utility function for money. Here, we also propose various types of linearity, adjusted to our special decision setting with subsets.

The first kind of linearity we might want to impose is on the transformation ϕ. Since it measures cardinal attitudes on the space of lotteries, it can be interpreted as a measure of second order risk aversion. In other words, if ϕ really is nonlinear, then the agent might treat compound lotteries differently than their reductions. So we impose the following.

Axiom 9 (Linearity). There exists a representation (ϕ, u, μ) such that ϕ is linear.[19]

By itself, linearity of ϕ does not identify a single preference, because μ could be one of many possible measures. We need more axioms to identify an ambiguity neutral probability assessment.

[19] The axiom might be objectionable because it is really defined on the utility rather than on the preference. We could move towards explicitly defining preferences over compound lotteries, but that would involve expanding our original domain of preference.

To proceed, we introduce linear operations in the choice space. Define $\alpha A + (1-\alpha)B = \{\alpha a + (1-\alpha)b : a \in A, b \in B\}$. The definition of singleton independence can now be generalized to include an ambiguous set on one side.

Axiom 10 (Singleton-set independence). $\{a\} \succsim A$ if and only if $\alpha\{a\} + (1-\alpha)\{x\} \succsim \alpha A + (1-\alpha)\{x\}$ for all $x \in \Delta X$, $\alpha \in (0,1)$.

The axiom states that the agent is neutral to mixtures between a set and a singleton. This is similar to the standard "hedging" axiom in subjective theories of ambiguity. The spirit of these axioms is that the agent prefers a mixture of two acts to either of the acts separately; the mixture "hedges" the ambiguity of the acts. For example, in Gilboa-Schmeidler (1989), the key axiom characterizing ambiguity aversion is that for all acts $f \sim g$, $\alpha f + (1-\alpha)g \succsim f$, while $\alpha f + (1-\alpha)g \sim f$ characterizes ambiguity neutrality. Our condition is that if the agent is indifferent between unambiguous set a and an ambiguous set A, then she will retain that indifference when these choices are mixed with another unambiguous choice.

The traditional notion of independence between sets is $A \succsim B$ if and only if $\alpha A + (1-\alpha)C \succsim \alpha B + (1-\alpha C)$ for any set C. This stronger form is often assumed in studies over sets of lotteries, where it is tied to the temporal resolution of uncertainty. In this interpretation, the mixture $\alpha A + (1-\alpha)C$ is viewed as the lottery where the set A is realized with probability α and C is realized with probability $1-\alpha$. Our model precludes this view, since the relative probabilities of A and C are fixed by μ. The probabilistic interpretation of set mixtures violates the agent's own assessment of the relative probabilities of these sets, determined by $\mu(A)$ and $\mu(C)$. Indeed, it is easy to find examples of preferences that meet our axioms and violate set independence. Standard set independence is too strong for our purposes, so we use the weaker singleton-set independence, Axiom 10.

Another notion of linearity does not involve mixtures, but instead looks at jumps or shifts in the linear space. Let $A + y = \{a + y : a \in A\}$ for any signed measure y on ΔX. Then $A + y$ is a translation of A within ΔX if $y(\Delta X) = 0$, because any mass taken away from sets by y will be assigned to other sets, preserving total mass at unity.

Axiom 11 (Translation independence). $\{a\} \succsim A$ if and only if $a + y \succsim A + y$ for any y with $y(\Delta X) = 0$ such that $a + y, A + y \in \mathcal{K}^*$.

This property has intuitive geometric appeal. A linear shift or translation of a set yields exactly the same translation for its ambiguity-free equivalents.

Each of these two definitions captures a type of linearity. The former involves linear mixtures in our space, while the latter involves linear jumps. It turns out that the singleton-set independence is a stronger form of linearity than translation independence.

Theorem 7. *If $\succsim$ is singleton-set independent, then $\succsim$ is translation independent.*

Proof. Take $a \sim A$ and y with $a + y, A + y \in \mathcal{K}^*(\Delta X)$. Simple algebra verifies

$$(1-\alpha)A + \alpha(a+y) = \alpha[(1-\alpha)a + \alpha(a+y)] + (1-\alpha)[A + \alpha y].$$

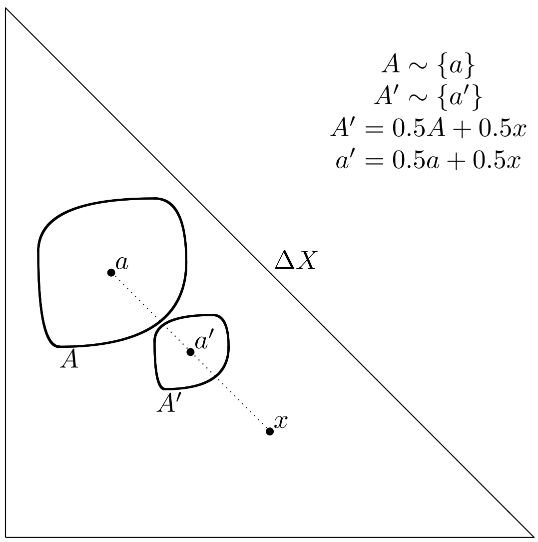

Figure 3: Singleton-set independence

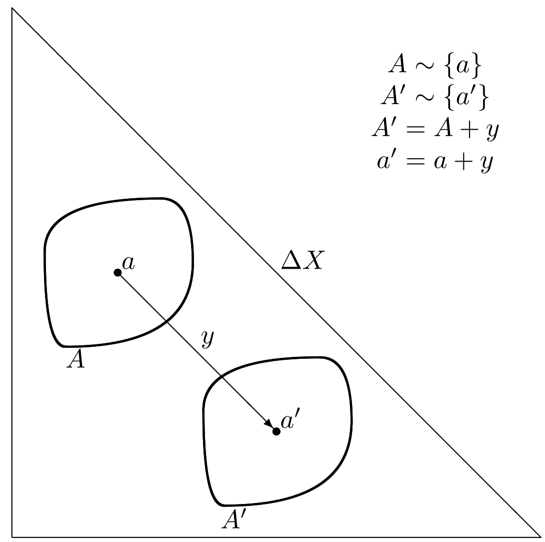

Figure 4: Translation independence

Figure 3 illustrates singleton-set independence. The singleton a is an ambiguity-free equivalent of the set A. A' is a mixture of A and the singleton x, $A' = \frac{1}{2}A + \frac{1}{2}x$. Accordingly, the ambiguity-free equivalent of A' is $a' = \frac{1}{2}a + \frac{1}{2}x$, the same mixture of A's equivalent and x. Figure 4 shows an example of translation independence. The set A is translated by the measure y to $A' = A + y$. Its ambiguity-free equivalent a also gets shifted by y to $a' = a + y$.

Then reflexivity implies

$$(1-\alpha)A + \alpha(a+y) \sim \alpha[(1-\alpha)a + \alpha(a+y)] + (1-\alpha)[A + \alpha y].$$

By singleton-set independence,

$$(1-\alpha)A + \alpha(a+y) \sim (1-\alpha)a + \alpha(a+y).$$

These two indifference relations and another application of singleton-set independence force

$$(1-\alpha)a + \alpha(a+y) \sim A + \alpha y.$$

Then, taking $\alpha \to 1$ and invoking continuity obtains $a + y \sim A + y$. The other direction is similar, expressing $A = (A+y) + (-y)$. □

We now provide a condition on the weighting measure μ that identifies it directly. If we take a measurable set E and shift it by y, then the shift $E + y$ should carry the same measure as E.

Definition 3. A probability measure $\mu \in \Delta^2 X$ is *translation invariant* if $\mu(E) = \mu(E+y)$ whenever $E \in \mathcal{B}(\Delta X)$, $y(\Delta X) = 0$, and $E + y \in \mathcal{B}(\Delta X)$.

Theorem 8. *There exists a unique regular translation invariant $\lambda \in \Delta^2 X$ such that $\mu(O) > 0$ for any open set O.*[20]

Proof. See Appendix A.2.

We call this unique measure the Haar probability measure on ΔX and reserve the symbol λ to denote this measure.[21] When X is finite with n elements, we can identify ΔX as

$$\Delta X \cong \{x \in \mathbb{R}^{n-1} : x_1, \ldots, x_{n-1} \geq 0, \sum_{i=1}^{n-1} x_i \leq 1\}$$

with the algebra of vector addition. Then λ is the Lebesgue measure in $\mathbb{R}^{n-1}$ restricted to ΔX.

We introduced two kinds of linearity on the preferences, singleton-set independence and translation invariance. Translation independence of the preference does not guarantee translation invariance of the measure μ; consider the space of lotteries on two consequences represented on the interval $[0, 1]$. Set ϕ as the identity function and μ as any exponential distribution. Then (ϕ, u, μ) is translation invariant, yet μ is not the uniform distribution. So, we require the stronger axiom of singleton-set independence, which does guarantee that μ is the Haar probability measure.

Theorem 9. *$\succsim$ meets Axioms 9 and 10 if and only if there exists a representation (ϕ, u, μ) where ϕ is the identity function and $\mu = \lambda$.*[22]

Proof. See Appendix A.4.

Corollary 10. *Suppose X is finite with n elements. $\succsim$ meets Axioms 9 and 10 if and only if there exists a representation (ϕ, u, μ) where ϕ is the identity function and μ is identified with the uniform distribution on the simplex $\Delta X \subset \mathbb{R}^{n-1}$.*

Since ambiguity neutrality is uniquely identified, we can now naturally define ambiguity aversion as any preference which is more ambiguity averse than the representation of Theorem 9. Table 1 shows the similarities between the classic elements of the theory of risk and the elements of our theory of ambiguity. We hope the related intuitions are conducive for applications.

6 Applications

6.1 Reinsurance and trade of uncertain prospects

Our first application is in the field of risk sharing or insurance. From the perspective of risk theory, the market for reinsurance is puzzling. Direct insurers usually purchase protection from a

[20] A measure μ is *outer regular* if $\mu(A) = \inf\{\mu(O) : A \subseteq O, O \text{ open}\}$. It is *inner regular* if $\mu(A) = \sup\{\mu(K) : K \subseteq A, K \text{ compact}\}$. It is *regular* if it is both inner and outer regular. Although this usage seems to be standard, some textbooks use the term differently.

[21] The term Haar measure is loosely adopted from the mathematics of measurable groups. Strictly speaking, λ is not a Haar measure; λ is actually a restriction of a Haar measure on a special topological group constructed in the proof of Theorem 8.

[22] Theorem 9 also provides decision theoretic foundations for the ignorance prior often used in Bayesian statistics.

	Risk	Ambiguity
Moving parts	v	ϕ, μ
Extreme	Wald criterion	Maxmin
Arrow–Pratt	$-v''/v'$	$-\phi''/\phi'$
Equivalent	$\delta_x \sim l$	$l \sim A$
More averse	$l \succsim_1 \delta_x \Rightarrow l \succsim_2 \delta_x$	$A \succsim_1 l \Rightarrow A \succsim_2 l$
Neutrality	Linear v	Linear ϕ, Haar μ

Table 1: Analogs between risk and ambiguity

reinsurance company against catastrophes; contracts often specify that the reinsurer absorbs any of the direct insurer's losses over a certain ceiling. If the reinsurer is really more risk acceptant than the primary insurance provider, then the optimal risk sharing scheme has the direct insurer cede *all* of its risk to the reinsurer. Furthermore, if the insurance company is publicly owned, this risk reduction is redundant given the diversification of its investors' private portfolios (Doherty and Tinic 1981). Therefore, institutional details beyond risk sharing, such as the tax code, are often invoked to explain reinsurance.[23] Even if these details help explain reinsurance against general excess losses, they cannot explain why reinsurance contracts often have contingent exclusions.[24]

If agents separate risk and ambiguity, then they will insure against these two types of uncertainty in different ways. Catastrophic losses are often rare outcomes. Without actuarial data, insurers cannot form precise beliefs about these outcomes. This is important if the direct insurer company is very willing to accept risk, but hesitant to accept ambiguity. On the other hand, reinsurers may be more ambiguity acceptant and willing to provide insurance against ambiguous uncertainties.[25] Finally, there are some uncertainties that are so ambiguous that even the reinsurer refuses to accept them, contingently excluding coverage for certain events.

Using the standard model of risk, two agents with similar risk profiles will never trade uncertain prospects. If one agent prefers the other's prospect, then that other agent also prefers to keep his prospect and rejects trade. To contrast, in our setting, two agents might have the same risk attitude, yet still trade bets because of differences in their ambiguity attitudes. We can go even further: it is possible for one agent to be more risk averse than another, yet engage in a trade of risky prospects

[23] See (Mayers and Smith 1990) for a survey and empirical analysis.

[24] For example, after a wave of urban riots in the late 1960's, reinsurance companies contractually excluded coverage against civil disorder. The federal government passed the Urban Property Protection and Reinsurance Act of 1968. This legislation established the Federal Insurance Administration, which offered reinsurance *specifically against riots*. More recently, after 11 September 2001, many reinsurers dropped coverage for terrorist attacks. The insurance industry has since lobbied the federal government to provide terrorism reinsurance. The British government already owns and administers Pool Re, a company that reinsures only against terrorism.

[25] A former vice president of a major reinsurance company mused: "Furthermore, the direct insurer, by the medium of reinsurance, tries to control his own books so that he has a fairly homogeneous group of risks retained, each for an amount that will permit the law of large numbers to work with considerable predictability. Compared to this, the reinsurer is something of a riverboat gambler. His stakes are higher and the odds are correspondingly longer (McCullough 1964)."

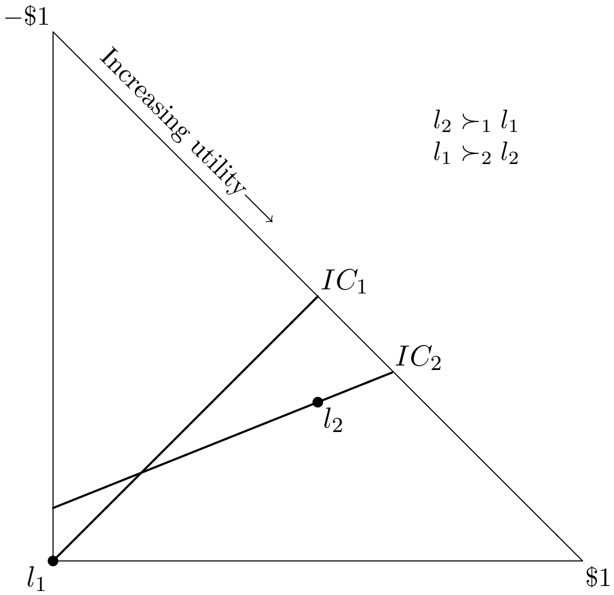

Figure 5: Insurance against risk

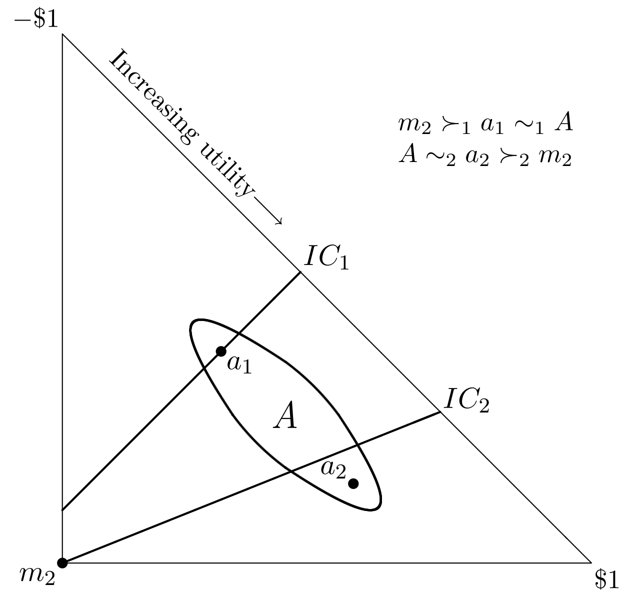

Figure 6: Insurance against ambiguity

Ann is risk neutral with indifference curves IC_1; Bob is risk averse with indifference curves IC_2. In Figure 5, Ann is endowed with the riskless l_1 and Bob is endowed with the risky l_2. Both prefer the other's endowment, so they trade positions: Ann provides Bob with insurance against risk. In Figure 6, Ann starts with the ambiguous set A, while Bob starts with the riskless and unambiguous lottery m_2. Ann's ambiguity-free equivalent for A is a_1; Bob's equivalent is a_2. Because their equivalents differ, they can trade: Bob provides Ann with insurance against ambiguity.

where the more risk averse party ends up with the set of more volatile lotteries.

We illustrate these points graphically. Consider a consequence space consisting of three elements, $X = \{-\$1, \$0, \$1\}$. ΔX is represented by the triangle depicted in Figure 5; the horizontal axis measures the probability of \$1 and the vertical axis measures the probability of $-\$1$. Risk tolerance is captured by the slope of the indifference curves: a slope of one is risk neutral, less than one is risk averse, and greater than one is risk loving. In all cases, utility increases as the lines move southeast. Treat the triangle like an Edgeworth box, where two agents, call them Ann and Bob, can trade their endowed lotteries. Figure 5 depicts Ann and Bob initially holding the singleton lotteries l_1 and l_2 respectively. Ann is risk neutral and Bob is risk averse, captured by the slopes of their indifference curves, IC_1 and IC_2. Here, each wants to trade, since the other's lottery is to the right of the indifference curve IC_j of her own lottery. In fact, Ann will provide full insurance for Bob, since l_1 is the riskless lottery that yields \$0 with certainty. Imagining a similar picture, it is easy to see that strictly Pareto improving trades are impossible if both agents share the same risk attitude.

Now consider the situation in Figure 6. Ann is endowed with the ambiguous set A. Bob is endowed with the unambiguous point lottery m_2, which is the degenerate lottery that yields \$0 with certainty. They both carry the same risk attitudes from Figure 5, but Ann is relatively ambiguity averse while Bob is ambiguity loving. Because of this, they have different ambiguity-free equivalents for Ann's endowment A: Ann is indifferent between a_1 and A, while Bob is indifferent between a_2 and A. So, the initial indifference curves pass through a_1 for Ann and m_2 for Bob, drawn as IC_1 and IC_2 in the diagram. Since m_2 is to the southeast of IC_1 for Ann, she strictly prefers a trade. Similarly, since a_2 is southeast of IC_2, Bob also strictly prefers a trade. Trade is Pareto improving, even though each element of A has more variance in outcomes than the singleton m_2. Here, Bob insures Ann against ambiguity, even though he is more risk averse. This is driven by their different ambiguity-free equivalents for A.

At the same time, both parties would never trade m_2 for any single lottery $a \in A$ in Figure 6. To see this, suppose the expected value of the lottery a is positive. In the picture, that means a is southeast of the 45 degree line emanating from the origin. Then Ann prefers to keep a, because m_2 has an expected value of zero and she is risk neutral. On the other hand, suppose the expected value of a is negative. Then a is northwest of the 45 degree line. Here, Ann wants to trade a for m_2. However, since Bob is more risk averse than Ann, he also strictly prefers m_2 to a and refuses the trade. So, any single trade of m_2 for any $a \in A$ will be rejected, yet both sides will agree to trade m_2 for the entire set A. Even if individual trade is impossible for any particular element of an ambiguous set, trade for the entire set might still be Pareto improving. We could construct a similar diagram where both Ann and Bob are risk neutral yet find a Pareto improving trade, even though such trade is impossible in the entire domain of singleton endowments.

Finally, notice that the story for risk neutral trade is impossible if we assume the maxmin representation of Theorem 1. Under risk neutrality, both Ann and Bob would share a common ambiguity-free equivalent for the set A, namely $\min_{a \in A} u(a)$. Then trade between ambiguous prospects decomposes into trade between shared ambiguity-free equivalents, which is captured by the standard theory on singletons.

6.2 Game theory

There has already been some research on games with ambiguous beliefs; the following discussion is most closely related to work by Klibanoff (1996) and Lo (1996). Both papers expand the notion of Nash equilibrium to allow for sets of beliefs over the other's strategies and assume that this ambiguity is resolved using a maxmin criterion, which we discussed in Section 3. We admit a more general class of utility functions on sets, axiomatically justified by Section 4.

We expand the domain of actions to include entire sets of mixed strategies, rather than single mixed strategies, in our solution concept. Accordingly, preferences are defined over sets of distributions over outcomes. An ambiguous two-player game G consists of strategies S_i, sure outcomes $S_1 \times S_2$, and ambiguity preferences $\succsim_i$ defined on $\mathcal{K}(\Delta(S_1 \times S_2))$.[26] Any two compact sets of mixed strategies $A_1 \in \Delta S_1, A_2 \in \Delta S_2$ generate a compact set of (product) distributions on $S_1 \times S_2$.

$$[A_1 \times A_2] = \{[\sigma_1 \times \sigma_2] : \sigma_1 \in A_1, \sigma_2 \in A_2\} \in \mathcal{K}(\Delta(S_1 \times S_2)).^{27}$$

[26] We limit ourselves to two players for expositional ease. If S_i is uncountable, there is an understood σ-algebra.
[27] Here, $[\sigma_1 \times \sigma_2]$ refers to the product measure of σ_1 and σ_2 on $S_1 \times S_2$.

Our solution concept incorporates *sets* of mixed strategies. It is a simple generalization of the "Beliefs Equilibrium with Agreement" introduced by Lo (1996) that allows for a richer resolution of ambiguity.

Definition 4. Two compact sets of mixed strategies $A_1^* \subseteq \Delta S_1, A_2^* \subseteq \Delta S_2$ comprise an *ambiguous (mixed) equilibrium* if for $i = 1, 2$,

$$[\sigma_i^* \times A_{-i}^*] \succsim_i [\sigma_i \times A_{-i}^*]$$

for all $\sigma_i^* \in A_i^*$, $\sigma_i \in \Delta S$.

In our equilibrium, an agent may leave his choice of strategy ambiguous, as to induce the other player to select certain strategies in response. If players are responsive to ambiguity over X, it makes sense that they would remain ambiguous about their own strategy to manipulate this response. Equilibrium requires that every element of A_i^* is still a best response to the set of potential strategies by the other player.[28]

The interpretation of equilibrium as a profile of beliefs or conjectures on the other's strategy is especially appropriate in the context of ambiguity. In this view, the agent actually plays some strategy, but this strategy is supported by an uncertain conjecture about the other's strategy. In an ambiguous equilibrium, this conjecture involves both risk and ambiguity.[29]

Observe that any Nash equilibrium is an ambiguous equilibrium. The following is an immediate consequence of the classic existence theorem (Nash 1950).

Theorem 11. *Suppose for $i = 1, 2$, S_i is finite and $\succsim_i$ meets singleton independence. Then there exists an ambiguous equilibrium.*

6.3 Mean–variance–ambiguity frontier

We derive an analog to the classic mean–variance frontier of financial portfolios. The analysis does not appeal to any state space but only on an ambiguous conjecture on the distribution of joint returns. Consider a finite number of assets. The joint return on these assets draws from a set of multivariable normal distributions with possible means $\mathbf{e}(t)$ and common variance-covariance matrix $\mathbf{V}$:

$$\{\mathcal{N}[\mathbf{e}(t), \mathbf{V}] : t \in [0,1]\}.^{30}$$

We assume that for $\mathbf{e} : [0,1] \to \mathbb{R}^n$ each i component $e_i : [0,1] \to \mathbb{R}$, is increasing. This means we can view t as capturing some sort of fundamental uncertainty, like the general state of the economy or the interest rate, that determines whether the expected return on the market will be generally high or low.

[28] The definition compares singleton strategies for the player's own actions. This precludes self-deception: an agent knows which mixture she will play. The definition could be naturally weakend to allow self-deception.

[29] For a more involved discussion, see (Lo 1996).

[30] We assume each asset carries some risk; the analysis with a single riskless asset is similar, but more algebraically cumbersome.

Each stock is then characterized by a set of marginal distributions, with a lowest possible expected return of $e_i(0)$ and a highest possible expected return of $e_i(1)$. Let e_i refer to the constant $(e_i(1)+e_i(0))/2$ and $d_i = e_i(1) - e_i$. Then d_i is a measure of the dispersion of the mean for asset i, which we associate with the ambiguity of the asset. These parameters are illustrated in Figure 7. Let $\mathbf{e} = [e_1 \cdots e_N]^\top$ and $\mathbf{d} = [d_1 \cdots d_N]^\top$. A portfolio is a vector of weights $\mathbf{w}$ on the different assets with $\sum_i w_i = 1$.[31]

Like the expected return, the dispersion of a portfolio $\mathbf{w}$ is equal to the sum of its weighted component dispersions, $\sum_i w_i d_i$. This is justified by our parameterization of the ambiguity on the joint returns by some underlying ambiguity t.[32]

The minimum variance portfolio $\mathbf{w}^*(\mu, \nu)$ with mean μ and dispersion ν solves:

$$\min \frac{1}{2}\mathbf{w}^\top \mathbf{V}\mathbf{w}$$

subject to

$$\begin{aligned} \mathbf{w}^\top \mathbf{e} &= \mu, \\ \mathbf{w}^\top \mathbf{d} &= \nu, \\ \mathbf{w}^\top \mathbf{1} &= 1, \end{aligned}$$

where $\mathbf{1}$ is the n-vector of ones.[33]

Because we are able to formulate ambiguity in the form of linear dispersion, it is simple to generalize the classic results from the standard analysis of the mean–variance frontier. First, the optimal portfolio weights are linear in the required mean return and dispersion.

Proposition 12. *The minimum variance portfolio $\mathbf{w}^*(\mu, \nu)$ is linear in μ and ν.*

Proof. All proofs from this section are in Appendix A.5.

The classic mean–variance frontier is shaped as a parabola in two dimensions. Our frontier lives in three dimensions, and is shaped as a paraboloid. If the agent prefers less ambiguity, the quadrant of the paraboloid that has high mean and low ambiguity is the efficient region of the frontier. If we fix a level of ambiguity, the resulting ν slice is the classic parabola frontier.

Proposition 13. *The mean–variance–ambiguity frontier is a paraboloid.*

Finally, the standard analysis yields the classic two fund separation result. Here, we achieve three fund separation. Any frontier portfolio can be achieved by buying positions in three mutual

[31] The ambiguity here is not over all distributions ΔX over wealth, but over a parameterized family of normal distributions. This still falls in the purview of the theory, as explained by the remark on page 13.

[32] While the structure of the summed dispersion simplifies the algebra, we could make other assumptions on the aggregation of ambiguity. For example, we could assume that the dispersion of a portfolio is $\sum_i |w_i| \cdot d_i$. This is justified if the set of joint means is rectangular, and differs from the assumption in the text only in the presence of short sales. The analysis is similar but more tedious: a variety of Kuhn–Tucker conditions must be checked. In any case, all of the following discussion would still hold "locally."

[33] The constraints are feasible if there are at least three assets and $\mathbf{e}, \mathbf{d}, \mathbf{1}$ are linearly independent.

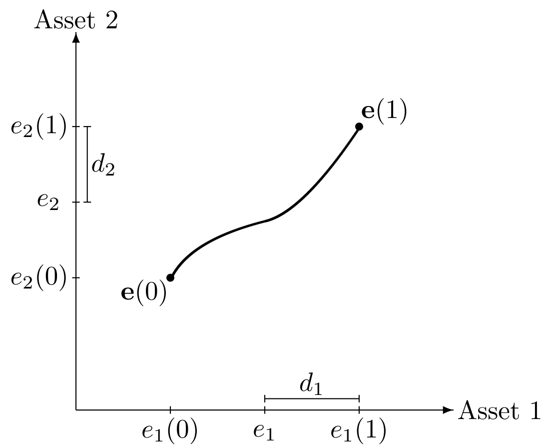

Figure 7: Joint means for ambiguous financial assets

The figure illustrates the parameters used in constructing the mean–variance–ambiguity frontier. The parametrized curve $\mathbf{e}$ shows the possible joint expected returns. The point $e_i(1)$ is the best possible expected return for asset i, $e_i(0)$ is the worst. The median $e_i = [e_i(0) + e_i(1)]/2$. The disperstion $d_i = e_i(1) - e_i$ reflects the ambiguity of asset i.

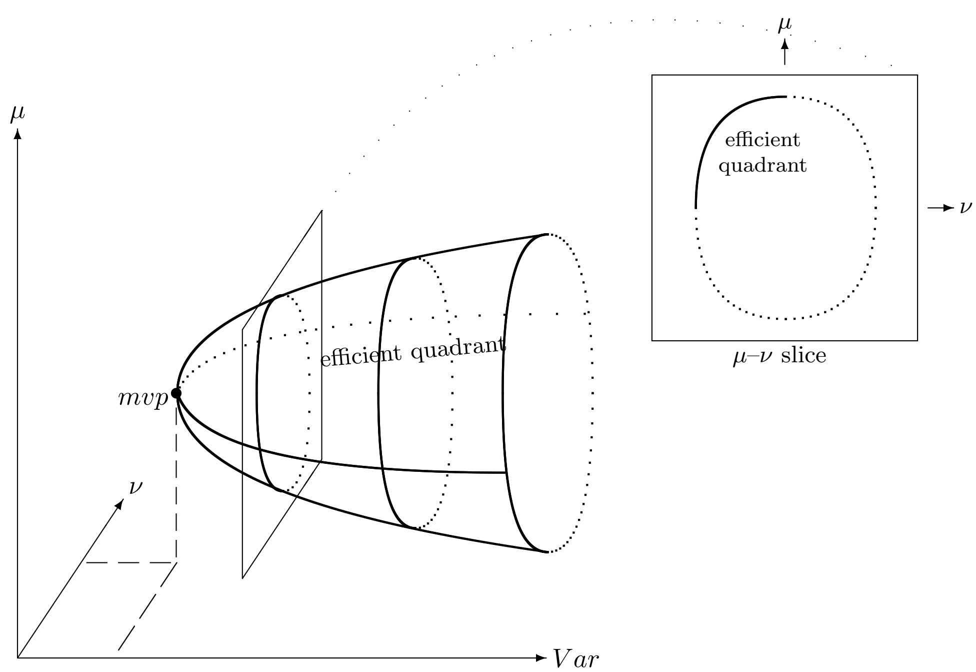

Figure 8: Mean–variance–ambiguity frontier

The paraboloid is the mean–variance–ambiguity frontier. The point mvp is the globally minimum variance portfolio. The quadrant closest to the reader and towards the top is the set of efficient portfolios when the agent is risk and ambiguity averse. This is perhaps clearer from the inset, which shows a slice of the frontier in the mean (μ) and ambiguity (ν) dimensions.

funds on the frontier. The additional fund is necessary to capture different dispersions among the assets.

Proposition 14. *The mean–variance–ambiguity frontier is spanned by three portfolios.*

7 Concluding remarks

In this paper, we suggest an alternative formulation for decisions under ambiguity: sets of lotteries over consequences. We characterize the following utility for a set A:

$$U(A) = \frac{\int_A \phi \circ u \, d\mu}{\mu(A)},$$

where u is an expected utility on lotteries, ϕ is an increasing transformation, and μ is a second order measure on lotteries. This representation offers two novel devices for capturing ambiguity aversion: the curvature of the transformation ϕ and the weighting of the measure μ. We provide a definition of comparative ambiguity aversion and a unique characterization of absolute ambiguity neutrality; these are conceptually linked to similar notions in the theory of risk. The decision maker becomes more ambiguity averse as ϕ becomes more concave or as μ becomes more pessimistic; she approaches maxmin expected utility at the limit.

The main contribution of the theory may be the domain of choice itself, which incorporates ambiguity without invoking an explicit state space. We believe that this framework is closer to the cognitive protocol used by actual decision makers in resolving complex decisions, where the probabilities are imprecise and the states are intractable. Our approach, which questions the foundations of the Savage formulation, runs philosophically counter to the usual approach to ambiguity, which introduces a more involved utility function or a more explicitly specified state space within the basic Savage structure. Yet the descriptive shortcomings of subjective expected utility may not be in a particular axiom, but in the primitives of the model. While the literature currently allows for a variety of axioms within the vNM or Savage schemes, perhaps a more catholic view of basic formulations is also appropriate.

A Appendix

A.1 Proof of Theorem 1

We omit the straightforward proof of necessity and proceed to sufficiency. By (Parthasarathy 1967, Theorem 6.2), ΔX inherits separability from X and a simple mixture argument shows ΔX is connected. This allows us to invoke (Debreu 1954, Theorem 1): continuity implies there exists some continuous $u : \Delta X \to \mathbb{R}$ such that u represents $\succsim$ restricted to the singletons. Define $U(A) = \min_{a \in A} u(a)$, which is well defined as u is continuous and A is compact. Pick A, B such that $U(A) \geq U(B)$. Let a^*, b^* refer to the respective minimizers. $\{a^*\} \succsim \{b^*\}$, by our construction of U and u. $\{b^*\} \succsim B$ as $\succsim$ is decreasing. Thus $\{a^*\} \succsim B$. Now consider a sequence of finite unions $A_n = \bigcup_{i=1}^n \{a_i\}$ with $a_i \in A$ such that $A_n \to A$ in the Hausdorff metric. Such a sequence exists because finite sets are dense in the Vietoris topology on the nonempty subsets of ΔX when ΔX is Polish.[34]. Finite sets are thus dense in the Hausdorff metric topology, which is the relativization of the Vietoris topology to compact sets. $A_n \succsim \{a^*\}$ since $\{a_i\} \succsim \{a^*\}$ as a^* minimizes u on A and iteratively applying disjoint upper closure. Then $A \succsim \{a^*\} \succsim B$ by continuity.

Next, suppose $A \succsim B$ with u-minimal elements a^*, b^*. Then $\{a^*\} \succsim A$ by decreasingness. $B \succsim \{b^*\}$ using a finite approximation argument similar to the one above. Thus $\{a^*\} \succsim A \succsim B \succsim \{b^*\}$, implying $u(a^*) \geq u(b^*)$. □

A.2 Proof of Theorem 8

Contrary to the order of presentation, we prove Theorem 8 before Theorem 3. We will use the measure constructed here in proving Theorem 3. The proof depends on the theory of Haar measure.

Definition 5. A Borel measure μ on a topological abelian group T is a *Haar measure* if $\mu(O) > 0$ for every nonempty open set O and $\mu(Et) = E$ for every Borel set E and $t \in T$.[35]

Under certain topological conditions, a Haar measure is guaranteed to exist. The following is (Halmos 1974, Theorem B, p. 254).

Theorem 15 (Haar). *In every locally compact [Hausdorff] topological group T, there exists at least one regular Haar measure.*[36]

In addition, the Haar measure is unique up to scale transformations (Halmos 1974, p. 263).

Theorem 16 (Haar). *If μ and ν are regular Haar measures on a locally compact [Hausdorff] topological group T, then there exists a positive finite constant c such that $\mu(A) = c\nu(A)$ for every Borel set E.*

Of course, ΔX is not a group, since it is not closed under the natural binary operation of adding measures pointwise. We will build an artificial topological group and an associated Haar measure. The probability measure required in the theorem is a restriction of a Haar measure to ΔX.

Let $ca(X)$ refer to the space of all finite Borel (signed) measures on X. For any integer k, let

$$T_k = \{m \in ca(X) : m(X) = k\}.$$

Endow each T_n with the topology of weak convergence, which we denote τ_k. Let

$$T = \bigcup_k T_k.$$

T is endowed with the topology generated by $\bigcup_k \tau_k$, which is equivalent to its relative topology in $ca(X)$ with the Prohorov metric.

Remark. The choice of T and its topology is delicate. The space $ca(X)$ with the topology of weak convergence may seem more natural, but presents two problems. First, $ca(X)$ is a Hausdorff topological vector space. Any locally compact Hausdorff topological vector space is finite-dimensional (Aliprantis and Border 1999,

[34]For a definition of the Vietoris topology, see (Klein and Thompson 1984, pp. 7–8)

[35]An abelian group T carrying a topology is a *topological group* if the inverse translation function $t \mapsto t^{-1}s$ is continuous for every $s \in T$.

[36]T is *locally compact* if every $t \in T$ has a compact neighborhood. T is *Hausdorff* if any two distinct points have disjoint neighborhoods.

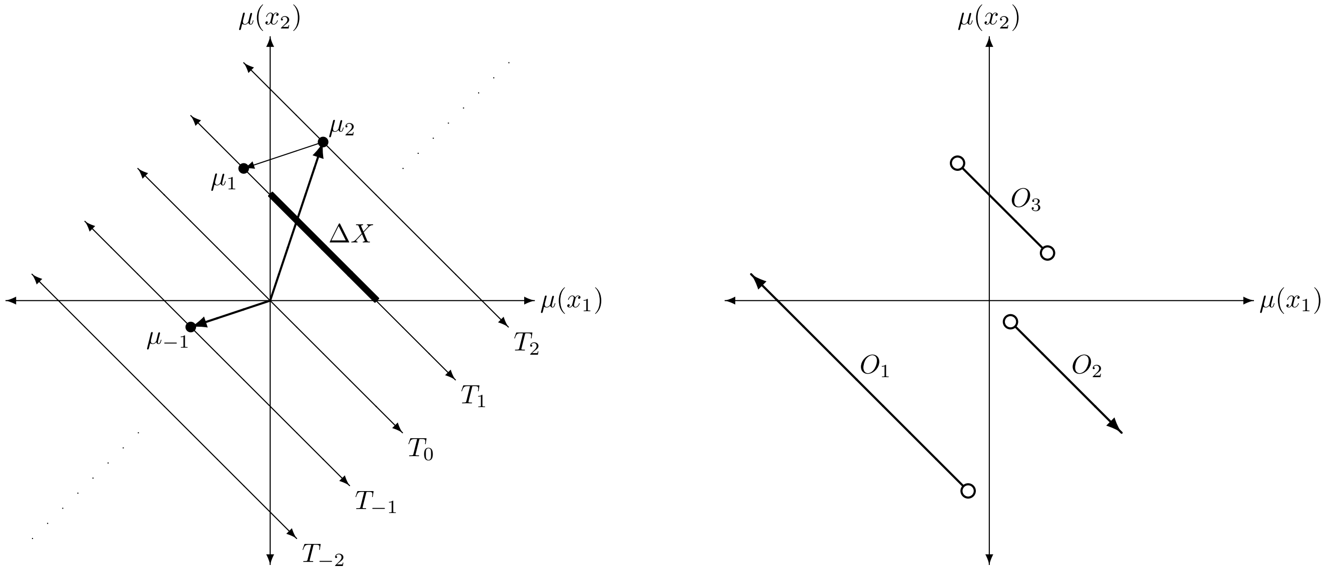

Figure 9: Constructions in the proof of Theorem 8

This figure illustrates the constructions for the special case $X = \{x_1, x_2\}$, where the set of measures $ca(X)$ can be represented as $\mathbb{R}^2$. On the left is the topological group T, which is the union of the lines T_k. T is closed under addition, as demonstrated in the picture by $\mu_1 = \mu_2 + \mu_{-1}$. The heavy line is $\Delta X \subset T_1 \subset T$. On the right, the three line segments O_1, O_2, and O_3, along with their unions, are (relatively) open sets in the topology on T. These sets are *not* open in the standard topology.

Theorem 5.69), while $ca(X)$ is infinite-dimensional whenever X is infinite. Then $ca(X)$ cannot be locally compact. Unfortunately, the existence of Haar measure is guaranteed only in locally compact topologies. Second, ΔX is a closed set with empty interior in this topology. Even if the Haar measure did exist, it would vanish at ΔX. Then speaking of its restriction to ΔX is senseless.

Lemma 17. *T is a topological abelian group under the operation $+$.*

Proof. It is easy to verify that T is an abelian group with the zero measure as its identity and $-m$ as the inverse of m. Continuity of inverse translation follows from the definition of weak convergence and the additivity of the integral: $\int f\, d[m - m'] = \int f\, dm - \int f\, dm'$. □

Lemma 18. *T is locally compact.*

Proof. Since X is compact, ΔX is compact. Fix $l^* \in \text{int}(\Delta X)$. Let $C_\varepsilon(l^*)$ denote the closure of the ε–ball about l^* in T. $C_\varepsilon(l^*)$ is contained in $\Delta^2 X$ for sufficiently small $\varepsilon > 0$. Since ΔX is compact, its closed subsets are compact. Therefore, $C_\varepsilon(l^*)$ is a compact neighborhood of l^* in T, considering the construction of its topology as the union of its relative component topologies on T_n.

Take any $m \in T$. Consider the function $f : C_\varepsilon(l^*) \to T$ defined by

$$f(l) = m + (l^* - l).$$

f is 1–1 and takes $C_\varepsilon(l^*)$ to its $m + l^*$ inverse translations, and is therefore continuous. Then $g : C_\varepsilon(l^*) \to f(C_\varepsilon(l^*))$ defined by $g(l) = f(l)$ is a homeomorphism: it is obviously onto and its inverse is also a translation, therefore continuous. Then g preserves nonempty interior and compactness, so $g(C_\varepsilon(l^*))$ also has nonempty interior in T and is compact, hence is a compact neighborhood of m. □

The preceding two Lemmas and Theorem 15 imply there exists a regular Haar measure on T. Any Haar measure is finite on compact sets, so we can pick the scaling such that $\lambda(\Delta X) = 1$. Restricting λ to the Borel subsets of ΔX provides the required translation invariant element of $\Delta^2 X$.

To prove uniqueness, observe that any translation invariant measure μ on ΔX can be extended to T. This is by first setting

$$\lambda_U(C) = \mu(A)\frac{C : U}{A : U}$$

for any open set U, nonempty regular set $A \subseteq \Delta X$, and compact C in T, where the ratio $E : F$ refers to the smallest integer n such that E can be covered by n translations of F. By taking the limit of sufficiently small U and invoking the Tychonoff Product Theorem, λ_U can be verified to be a content; this content can be extended to a measure using the Caratheodory extension procedure.[37] If μ is translation invariant on ΔX, its extension also is translation invariant. Then, by Theorem 16, there exists at most one translation invariant regular probability measure on ΔX. □

A.3 Proof of Theorem 3

We begin by proving a useful technical lemma. The lemma is used in the proof of Theorem 9 in Appendix A.4 and makes Corollary 4 an immediate consequence of Theorem 3.

Lemma 19. *Let $f = \frac{d\mu}{d\lambda}$, the Radon–Nikodym derivative of μ with respect to λ. f is continuous at a if and only if $\frac{\mu(A_n)}{\lambda(A_n)} \to f(x)$ whenever $A_n \to \{a\}$ in the Hausdorff metric topology and $\lambda(A_n) > 0$.*

Proof. We first prove sufficiency by contradiction. Suppose $\frac{\mu(A_n)}{\lambda(A_n)} \to f(a)$ if $A_n \to \{a\}$ and $f(x)$ is discontinuous at a. Then either $\max\{f(x), f(a)\}$ or $\min\{f(x), f(a)\}$ is discontinuous at a; assume without loss of generality it is the former. Then there exists some $\varepsilon > 0$ such that

$$D_n = \left\{x : \rho(x,a) < \frac{1}{n} \text{ and } f(a) - f(x) > \varepsilon\right\}$$

is nonempty for all n, recalling ρ is the Prohorov metric. Furthermore, $\lambda(D_n) > 0$ for all n otherwise we can find a function $g = f$ almost everywhere with respect to λ with g continuous at a. Then

$$\frac{\int_{D_n} f \, d\lambda}{\lambda(D_n)} < f(a) - \varepsilon$$

for all D_n while $D_n \to \{a\}$. This contradicts our assumption that $\frac{\mu(A_n)}{\lambda(A_n)} \to f(a)$ whenever $A_n \to \{a\}$.

For necessity, suppose f is continuous at a and take any sequence $A_n \to \{a\}$. Fix $\varepsilon > 0$. There is a corresponding $\delta > 0$ such that $|f(x) - f(a)| < \varepsilon$ whenever $\rho(x,a) < \delta$. As $A_n \to \{a\}$, there exists some N such that if $n > N$, $\rho(x,a) < \delta$ for any $x \in A_n$. Then for all $n > N$,

$$\begin{aligned}
\left|\frac{\mu(A_n)}{\lambda(A_n)} - f(a)\right| &= \left|\frac{\int_{A_n} f \, d\lambda}{\lambda(A_n)} - f(a)\right| \\
&\leq \frac{\int_{A_n} |f(x) - f(a)| \, d\lambda}{\lambda(A_n)} \\
&< \frac{\int_{A_n} \varepsilon \, d\lambda}{\lambda(A_n)} \\
&= \varepsilon.
\end{aligned}$$

[37] Details are similar to the argument in (Halmos 1974, p. 254–256).

□

Verifying the necessity of the axioms is routine and omitted. To prove sufficiency, we extend $\succsim$ to a larger family of sets. Let λ refer to the Haar measure on ΔX, which was discussed in detail in Section 5.2. We will consider the family of Borel sets modulo λ. Details of the following constructions can be found in (Halmos 1974, pp. 166–169). We write symmetric set difference as $A\Delta B = (A \setminus B) \cup (B \setminus A)$. Let $\mathcal{Z}'$ refer to the quotient group $\mathcal{B}(\Delta X)/\mathcal{N}$, where $\mathcal{N}$ is the subgroup of Borel sets with Haar measure zero. So, if two sets differ only on a set of Haar measure zero, $\lambda(A\Delta B) = 0$, then they are considered part of the same equivalence class in $\mathcal{Z}'$. We will remove the equivalence class of the empty set, to make $\mathcal{Z} = \mathcal{Z}' \setminus [\emptyset]$. At times, we will slightly abuse notation and denote the equivalence class $[A]$ by a representative element A; this should not cause any confusion.

We now provide a topology on $\mathcal{Z}$. Let

$$\pi(A, B) = \lambda(A\Delta B).$$

π is a metric on $\mathcal{Z}$, and is separable by (Halmos 1974, Theorem B, p. 168).

Lemma 20. *There exists an extension of $\succsim$ to $\mathcal{Z}$ that meets Axioms 5–8.*

Proof. The Haar measure λ is outer regular by definition. Then $[\mathcal{K}^*]$ is π-dense in $\mathcal{Z}$.[38] By the classic representation result (Debreu 1954, Theorem 1), there exists a continuous representation u on $\mathcal{K}^*$. The Hausdorff metric (d) topology on $\mathcal{K}^*$ is finer than π, since $K_n \xrightarrow{d} K$ only if $\bigcap_{n=0}^{\infty}(\mathcal{K}_n \Delta K) = \emptyset$. Then extend u to $\mathcal{Z}$ by setting

$$u(A) = \lim u(K_n)$$

for any increasing d-convergent sequence $K_0, K_1, \ldots \in \mathcal{K}^*$ such that

$$\lambda\left(\left[\bigcup_{n=0}^{\infty} \operatorname{int}(K_n)\right] \Delta A\right) = 0.$$

$\bigcup_{n=0}^{\infty} \operatorname{int}(K_n)$ is open. So $K = \overline{\bigcup_{n=0}^{\infty} \operatorname{int}(K_n)}$ is regular and has the same Haar measure as its interior. Therefore K is a regular set in $[A]$.

Meanwhile, $u(K) = \lim u(K_n) = u(A)$ by continuity of the original $\succsim$. Thus, for any $A \in \mathcal{Z}$, there exists some $K \in \mathcal{K}^*$ such that $u(K^*) = u(A)$ and $[K] = [A]$. Also, if $K \neq K'$ for regular sets K, K', then $K\Delta K'$ has a nonempty interior, implying $[K] \neq [K']$. So, the $K \in [A]$ found above is unique. Then preference ordering induced by the extended u extends $\succsim$ to $\mathcal{Z}$ and immediately inherits all of the axioms through the appropriate $K \in [A]$. It inherits continuity in the following manner:

$$d([A], [B]) = d(K_A, K_B),$$

for the unique regular compact $K_A \in [A]$, $K_B \in [B]$. □

Adding the singletons to $\mathcal{Z}$, when necessary, is done in the natural fashion. At times, we will move between the spaces with and without the singletons appended, but this should not cause any confusion.

We will prove that the axioms on the extended preference are sufficient for the utility representation on all of $\mathcal{Z}$. Then the utility will also represent the original $\succsim$ on the restricted domain $\mathcal{K}^*$.

[38] As is standard, by $[\mathcal{K}^*]$ we mean $\{[A] : A \in \mathcal{K}^*\}$.

To prove sufficiency, we reference the theory of probability representation on λ-systems of subsets. This theory was developed conceptually in the mathematical foundations of quantum mechanics (Birkhoff and von Neumann 1936, von Neumann 1955). An axiomatic development in that context can be found in (Suppes 1966), where the term "quantum mechanical algebra" replaces λ-system. More recently, such structure is exploited by Zhang (1999) and Epstein and Zhang (2001) to define the algebraic properties of unambiguous events in the Savage state space.

Definition 6. A family λ of subsets of X is a λ-*system* if:

1. $X \in \lambda$,
2. $S \in \lambda$ implies $S^{\complement} \in \lambda$, and
3. if $A_1, A_2, \ldots \in \lambda$, then $\bigcup_{n=1}^{\infty} A_n \in \lambda$.

The family of indifferent subsets of A, notated as λ_A, is conveniently a λ-system.

Lemma 21. *The family $\lambda_A = \{S \subseteq A : S \sim A\}$ is a λ-system.*

Proof. The first holds follows since the preference is reflexive. If $S \in \lambda_A$, then its (relative) complement $A \setminus S \sim A$ or we contradict disjoint set betweenness. Closure under finite disjoint unions follows directly from disjoint set betweenness. For any countable disjoint sequence $\{A_n\}_{n=1}^{\infty}$, $\lambda(\bigcup_{n=1}^{N} A_n)$ converges to $\lambda(\bigcup_{n=1}^{\infty} A_n)$ as N goes to infinity because $\lambda(\bigcup_{n=1}^{\infty} A_n)$ is finite. Then the result holds by applying disjoint set betweenness to the finite unions $\bigcup_{n=1}^{N} A_n$, then passing the limit to the preferences using continuity. $\square$

There are various results that find sufficient conditions on a qualitative likelihood ranking $\succeq_l$ on a λ-system for the existence of a consistent probability measure, for example (Suppes 1966, Theorem 3) or (Krantz, Luce, Suppes, and Tversky 1971, p. 215). We state a recent version by Zhang (1999), as reported in (Epstein and Zhang 2001, Theorem B.1).

Theorem 22 (Zhang). *There exists a unique finitely additive, convex-ranged probability measure P on λ such that $A \succeq_l B \Leftrightarrow P(A) \geq P(B)$ for all $A, B \in \lambda$ if and only if $\succeq_l$ satisfies:*

1. *$A \succeq_l \emptyset$ for any $A \in \lambda$.*
2. *$X \succeq_l \emptyset$.*
3. *$\succeq_l$ is a weak order.*
4. *If $A, B, C \in \lambda$ and $A \cap C = B \cap C = \emptyset$, then $A \succ_l B$ if and only if $A \cup C \succ_l B \cup C$.*
5. *For any two uniform partitions $\{A_i\}_{i=1}^{n}$ and $\{B_i\}_{i=1}^{n}$ of S in λ, $\bigcup_{i \in I} A_i \sim_l \bigcup_{i \in J} B_j$ if $|I| = |J|$.*
6. (a) *If $A \in \lambda$ and $A \succ_l \emptyset$, there is a finite partition $\{A_1, \ldots, A_n\}$ of X in λ such that:*
      i. *$A_i \subseteq A$ or $A_i \subseteq A^{\complement}$ for all A_i, and*
      ii. *$A \succ_l A_i$ for all A_i.*

   (b) *If $A, B, C \succ_l \emptyset$, $A \cup C = \emptyset$, and $B \succ_l A$, then there is a finite partition $\{C_1, \ldots, C_n\}$ of C in λ such that $B \succ_l A \cup C_i$ for all C_i.*
7. *If $\{A_n\}$ is a decreasing sequence in λ and $A^* \succ_l \bigcap_n A_n \succ_l A_*$ for some A^* and A_* in λ, then there exists N such that $A^* \succ_l A_*$ for all $n \geq N$.*

We can apply this theorem to λ_A and the resulting quantitative probability has some nice properties in terms of representing $\succsim$ on a restricted domain. Unlike some other characterizations of probability based on preference, we do not have to worry about removing null sets at the beginning. This is because disjoint set betweenness forces the decision maker to put some weight on all sets. We will later extend the domain of representation in two steps to the entire space.

Lemma 23. *Take any $A \succ B$ with $A \cap B = \emptyset$. There exist finitely additive, convex-ranged probability measures P_A and P_B on λ_A and λ_B such that:*

1. *$u(S \cup B) = P_A(S)$ is a utility representation of $\succsim$ on $\{S \cup B : S \in \lambda_A\}$; and*
2. *$u(A \cup T) = -P_B(T)$ is a utility representation of $\succsim$ on $\{A \cup T : T \in \lambda_B\}$.*

Furthermore, P_A is robust to choice of $B \prec A$ and P_B is robust to choice of $A \succ B$.

Proof. Define the likelihood ordering $\succeq_l$ on λ_A by $S_1 \succeq_l S_2$ if and only if $S_1 \cup B \succsim S_2 \cup B$. We now check the conditions in Theorem 22, omitting details. 1 and 2 are consequences of disjoint set betweenness. 3 follows since $\succsim$ is a weak order. 4 follows from balancedness. 5 follows from balancedness and disjoint set betweenness. 6 follows from divisibility. 7 follows from divisibility and continuity. We now have a convex-ranged probability measure P_A on λ_A. The utility $u(S \cup B) = P_A(S)$ meets the requirements of the lemma by the construction of P_A. Constructing the second measure P_B and utility is symmetric. The robustness claim follows from balancedness. □

Later in the proof, we will need to extend P_A to the entire level set. This step is not immediate because the family of all sets indifferent to A does not contain a superset to use as X. We now construct this extension.

Lemma 24. *There exists a finitely additive set function that extends P_A to the level set of A, which is unique up to a scale transformation.*

Proof. There exists a probability on any family of sets indifferent to each other. Fix $A \in \mathcal{Z}$. If $A \sim \Delta X$, then the extension is already provided by $P_{\Delta X}$. So, we may assume that $A \nsim \Delta X$. Now take any set B with $A \sim B$.

Case 1: $A \succ \Delta X$. Then pick $A^* \subset A$ and $B^* \subset B$ such that $A^* \cup B^* = \emptyset$, $P_A(A^*) = P_B(B^*)$, and $A^* \cup B^*$ is a strict subset of ΔX. The existence of such sets is guaranteed by the divisibility axiom. In view of disjoint set betweenness, we can find some $T \prec A$ which is disjoint from A^* and B^*. Assume without loss of generality that $A^* \cup T \succsim B^* \cup T$. Let $\alpha = \frac{P_A(A^*)}{P_B(B^*)}$. Then define $P_A(S) = \alpha P_B(S)$ for any $S \in \lambda_B$.

Case 2: $A \prec \Delta X$. Select A^* and B^* as in Case 1 and set $\beta = \frac{P_B(B^*)}{P_A(A^*)}$. Define $P_A(S) = \beta P_B(S)$ for any $A \in \lambda_B$.

Additivity of the extended P_A is a consequence of balancedness. P_A is the extension required in the Lemma, and its scaling depends on the choice of particular A in the indifference class of sets. □

Remark. Although the extended P_A is additive, it is *not* a measure. This is because the level set of A does not have a unique maximal element in the set containment order, which is implicitly required in the definition of a λ-system. The failure is not with the set function, but its domain.

The first application of Lemma 23 is in proving an intermediate preference result. In words, if A is preferred to B, we can "calibrate" preference by taking subsets of either the better set A or the worse set B. We will use this calibration result repeatedly in the proof.

Lemma 25. *Suppose $A \succsim B$ and $A \cup B = \emptyset$. If $A \succsim C \succsim A \cup B$, then there exists $B' \in \lambda_B$ such that $A \cup B' \sim C$. If $A \cup B \succsim C \succsim B$, then there exists $A' \in \lambda_A$ such that $A' \cup B \sim C$.*

Proof. If $A \sim B$, the result follows immediately from disjoint set-betweenness using A, B as A', B'. So take $A \succ B$ and assume the first case, $A \succ C \succ A \cup B$. Let P_B refer to the measure provided by Lemma 23 and the corresponding representation u. Since P_B is convex-valued, $P_B(\lambda_B) = [0,1]$, making u connected. Now, take any continuous utility representation f of $\succsim$ over all $\mathcal{Z}$, normalized so $f = u$ on $\{A \cup T : T \in \lambda_B\}$. Consider the π–continuous parameterization $g : [0,1] \to \mathcal{Z}$ defined by

$$g(\alpha) = [A \cup (\alpha B + (1-\alpha)\{b\})].$$

Recalling our topology, $[A] = [A \cup \{b\}] = [g(0)]$. Then $f \circ g$ is continuous in π with

$$f(A) = f(g(0)) \geq f(C) \geq f(g(1)) = f(A \cup B).$$

By the Intermediate Value Theorem, there exists some $x \in [0,1]$ with $f(g(x)) = f(C)$. Again, P_B is convex-valued. Thus, there exists some $B' \in \lambda_B$ with

$$u(A \cup B') = f(g(x)) = f(C),$$

proving the first statement of the lemma. The proof of the second statement is symmetric. □

In Lemma 23, we constructed a measure that represents preference on certain unions with respect to another fixed set. We now construct a measure that represents preference on all of $\mathcal{Z}$ with respect to another fixed set.

Lemma 26. *If $A \in \mathcal{Z}$, then there exists a nonatomic finitely additive signed measure ν_A on $\mathcal{Z}$ such that: $\nu_A(S) \geq (>) 0$ if and only if $S \succsim (\succ) A$.*

Proof. Since $\nu_a = \nu_A$ if $\{a\} = A$, we lose no generality by restricting attention to singletons. For a fixed $a \in \Delta X$, let

$$\bar{L} = \{l \in \Delta X : l \prec a\}$$

and

$$\mathcal{L} = \{L \in \mathcal{Z} : L \subseteq \bar{L}\}.$$

Similarly, let

$$\bar{U} = \{l \in \Delta X : l \succ a\}$$

and

$$\mathcal{U} = \{U \in \mathcal{Z} : U \subseteq \bar{U}\}.$$

$\mathcal{L}$ and $\mathcal{U}$ are respectively σ-algebras of $\bar{L}$ and $\bar{U}$, after appending the empty set. Either $\{a\} \succsim \Delta X$ or $\{a\} \precsim \Delta X$. We will assume the former; the argument for the second case is similar.

Fix any $U \in \mathcal{U}$. By our intermediate calibration result, Lemma 25, there exists some $L \in \lambda_{\bar{L}}$ such that $U \cup L \sim \{a\}$. Set

$$\nu(U) = P_{\bar{L}}(L),$$

where $P_{\bar{L}}$ is produced by Lemma 23. By the construction of $P_{\bar{L}}$, $\nu(U)$ is robust to our choice of L, hence uniquely defined. Select $U_1, U_2 \in \mathcal{Z}$ with $U_1 \cap U_2 = \emptyset$. There exists some $L_{12} \in \lambda_{\bar{L}}$ such that $L_{12} \cup U_1 \cup U_2 \sim$

$\{a\}$, by Lemma 25. Without loss of generality, assume $U_1 \succsim U_2$. By disjoint set betweenness, $L_{12} \cup U_1 \succsim \{a\}$. Then we can apply Lemma 25 again to L_{12} to find $L_1 \subset L_{12}$ with $L_1 \cup U_2 \sim \{a\}$. Set $L_2 = L_{12} \setminus L_1$. Another application of disjoint set betweenness forces $U_2 \cup L_2 \sim \{a\}$, since $L_1 \cup L_2 \cup U_1 \cup U_2 = (L_1 \cup U_1) \cup (L_2 \cup U_2) \sim \{a\}$ and $L_1 \cup U_1 \sim \{a\}$. Therefore, ν inherits disjoint additivity from $P_{\bar{L}}$. We now have a measure ν on $\mathcal{U}$.

Now take any $L \in \mathcal{L}$. If there exists some $U \in \mathcal{U}$ such that $L \cup U \sim \{a\}$, let

$$\nu(L) = -\nu(U).$$

On the other hand, if there is no $U \in \mathcal{U}$ with $L \cup U \sim \{a\}$, we can use Lemma 25 to produce a subset $L' \in \lambda_L$ such that there exists $U' \in \mathcal{U}$ with $L' \cup U' \sim \{a\}$. Let P_L refer to the measure on λ_L produced by Lemma 23. Set

$$\nu(L) = \frac{\nu(U')}{P_L(L')}.$$

The measure ν inherits disjoint additivity on $\mathcal{L}$ from $\mathcal{U}$ by its construction. This extends ν to $\mathcal{L}$.

We move to any arbitrary $A \subseteq \Delta X$. If $b \sim \{a\}$ for all $b \in A$, set $\nu(A) = 0$. Otherwise, we can express this set as $A = L \cup U$ for some $L \in \mathcal{L}$ and $, U \in \mathcal{U}$. Set

$$\nu(A) = \nu(L) + \nu(U).$$

Additivity is immediately inherited from $\mathcal{L}$ and $\mathcal{U}$. We have now extended ν to all of $\mathcal{Z}$.

Verifying the representation claim, suppose $A \succsim \{a\}$. Then $(S \cap \bar{L}) \cup (S \cap \bar{U}) \succsim \{a\}$. By Lemma 25, we can find a subset $U' \in \lambda_{S \cap \bar{U}}$ with $(S \cap \bar{L}) \cup U' \sim \{a\}$. Since $U' \subseteq S \cap \bar{U}$, $\nu(U') \leq \nu(S \cap \bar{U})$. Recalling the construction,

$$\begin{aligned} \nu(A) &= \nu(S \cap \bar{L}) + \nu(S \cap \bar{U}) \\ &= \nu(S \cap \bar{U}) - \nu(U') \\ &\geq 0. \end{aligned}$$

Similar arguments establish that $\nu(A) > 0$ only if $S \succ \{a\}$.

The measure $\nu(A)$ is convex-ranged, hence nonatomic. □

Up to this point, we have only considered finitely additive measures, sometimes called charges. We now show that they are also countably additive.

Lemma 27. *The measure ν_A is countably additive.*

Proof. ν_A is finite and ΔX is a Hausdorff space. Continuity implies the measures P_A of Lemma 23 are tight, which is passed on to ν_A by the construction of Lemma 26. Then (Aliprantis and Border 1999, Theorem 10.4) implies ν_A is countably additive. □

Now consider $ca(\mathcal{Z})$, the set of all countably additive finite (signed) measures on $\mathcal{Z}$. It is a Banach space once endowed with the total variation norm:

$$||\mu|| = \sup \left\{ \sum_{k=1}^{n} |\mu(A_k)| : \{A_1, \ldots, A_n\} \text{ is a partition of } \Delta X \right\}.$$

The representation features of ν_A are robust to positive scalar transformations, i.e. $\alpha \nu_A$ has the same properties whenever $\alpha > 0$. The measures constructed in Lemma 26 live in $ca(\mathcal{Z})$. The next result shows that

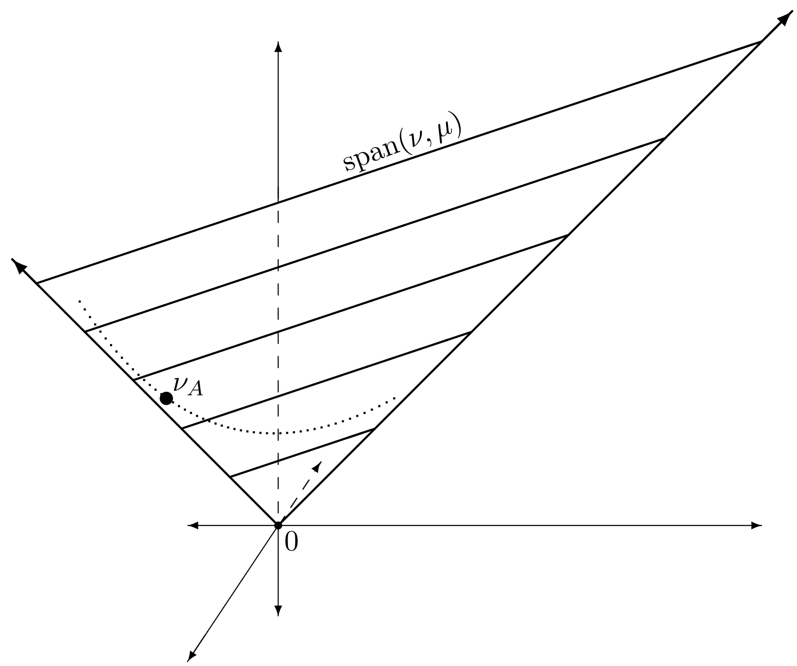

Figure 10: Lemma 28

This three-dimensional space is supposed to capture the infinite-dimensional space $ca(\mathcal{Z})$. Drawn with heavier lines, the plane cutting through the origin is the span of ν and μ. The measures ν_A, drawn as the dotted curve, live inside that span. A particular ν_A is labelled inside the curve.

these measures can be spanned by two elements of $ca(\mathcal{Z})$. This base will become the critical part of the representation.

Lemma 28. $\{\nu_A : A \in \mathcal{Z}\}$ *is spanned by measures* ν, μ *with* μ *a probability measure.*

Proof. Take any A, B, C which are not indifferent. We lose no generality by ordering them $A \succ B \succ C$. All the measures ν_A are nonatomic. We can invoke the Lyapunov Convexity Theorem: the image of the vector measure $[\nu_A, \nu_B, \nu_C]$,

$$[\nu_A, \nu_B, \nu_C](\mathcal{Z}) = \{(\nu_A(S), \nu_B(S), \nu_C(S)) \in \mathbb{R}^3 : S \in \mathcal{Z}\},$$

is convex.[39] Take any S with $\nu_A(S) = 0$. By construction, $\nu_B(S) > 0$ and $\nu_C(S) > 0$. By using Lemma 25, we can find $S^* \in \lambda_S$ with $S^* \cup L_1 \sim B$ and $S^* \cup L_2 \sim C$ for some L_1, L_2 disjoint from S^*. Recalling the representation condition in Lemma 26, $\nu_B(S^*) + \nu_B(L_1) = 0$ and $\nu_C(S^*) + \nu_C(L_1) = 0$. Now take any other T with $\nu_A(T) = 0$, and assume without loss of generality that $T \cap S^* = 0$. Then, suppose $\nu_B(T) \geq \nu_B(S^*)$. Then $\nu_C(T) \geq \nu_C(S^*)$, by disjoint set betweenness and the representation condition applied again. Therefore ν_B and ν_C induce the same ordering on $\{S : S \sim A\}$. Applying Lemma 24 and the uniqueness claim of Theorem 22, this ordering completely determines ν_B and ν_C on this restricted domain up to a scale transformation, i.e. $\nu_B(S) = c\nu_C(S)$ for a positive constant c, across any S with $\nu_A(S) = 0$. Then $\{x \in \mathbb{R}^3 : x_1 = 0\} \cap [\nu_A, \nu_B, \nu_C](\mathcal{Z})$ is contained in the ray $\{(0, ct, t) : t \geq 0\}$.

Since ν_A has strictly positive components (namely the strict upper contour set of A), the set

$$\{x \in [\nu_A, \nu_B, \nu_C](\mathcal{Z}) : x_1 > 0\}$$

[39] We thank Yossi Feinberg for referring the Lyapunov Convexity Theorem to us, which simplified an earlier proof.

is nonempty. Therefore, we can select vectors $x^0 \in [\nu_A, \nu_B, \nu_C](\mathcal{Z})$ such that $x^0_1 = 0$ and $x^1 \in [\nu_A, \nu_B, \nu_C](\mathcal{Z})$ such that $x^1_1 > 0$. We show x^0, x^1 span $[\nu_A, \nu_B, \nu_C](\mathcal{Z})$. Let $x^* \in [\nu_A, \nu_B, \nu_C](\mathcal{Z})$. We proceed by cases.

Case 1: $x^*_1 = 0$. Since $\{x \in \mathbb{R}^3 : x_1 = 0\} \cap [\nu_A, \nu_B, \nu_C](\mathcal{Z})$ is a ray, $x^* = cx^0$.

Case 2: $x^*_1 < 0$. Since $x^1_1 > 0$ is nonempty, we can find a convex combination $x' = \alpha x^* + (1-\alpha)x^1$ with $x'_1 = 0$. x' is in the vector range because it is a convex combination of two elements. But, applying Case 1 to x', we conclude x' is spanned by x^0. x^* is spanned by x' and x^1, while x' is a multiple of x^0. So x^* is spanned by x^0 and x^1.

Case 3: $x^*_1 > 0$. Apply Case 2 to $-x^*_1$.

Therefore there exist constants α, β such that $\nu_A = \alpha\nu_B + \beta\nu_C$. This suffices to show the entire space $\{\nu_A : A \in \mathcal{Z}\}$ can be spanned by any two of its measures, since the selection of A, B, C in the proof was arbitrary.

We finally show this span contains a strictly positive measure. Normalize each ν_A so $\nu_A(\bar{U}) = 1$, recalling $\bar{U} = \{l \in \Delta X : l \succsim A\}$. $\Delta(\mathcal{Z})$ is a closed subset of $ca(\mathcal{Z})$, while ν_A approaches $\Delta(\mathcal{Z})$ in the total variation norm as A approaches the $\succsim$–minimal element a_* in the order topology of $\succsim$. Since any subspace of $ca(\mathcal{Z})$ is closed in the total variation norm, the span contains a probability measure, which we can use as μ. □

Remark. Lemma 28 provides some insight into the uniqueness of u and μ in the utility function. Let $\nu_u(A) = \int_A u\, d\mu$. Alternative u' and μ' will also represent $\succsim$ if and only if the measures $\nu_{u'}$ and μ' span the same linear subspace of $ca(\mathcal{Z})$ as the measures ν_u and μ. So, uniqueness is identified linearly, but in a sense weaker than the affine families that identify the von Neumann–Morgenstern utility. This is partly because there are two moving parts, u and μ, in the representation.

Lemma 29. $\frac{\nu(A)}{\mu(A)}$ *is a utility representation.*

Proof. For fixed A, Lemma 28 guarantees coefficients $\alpha(A), \beta(A)$ such that

$$\alpha(A)\nu(A) - \beta(A)\mu(A) = 0 = \nu_A.$$

We demonstrate by contradiction that $\{\nu_A : A \in \mathcal{Z}\}$ is contained in a half space H with the zero measure on its boundary ∂H. If not, then there exist A, B and $\delta > 0$ with $\delta\nu_A = -\nu_B$. Take any set C with $C \succ A$ and $C \succ B$. By Lemma 26, $\nu_A(C) > 0$, implying $\nu_B(C) < 0$. This contradicts the initial assumption that $C \succ B$. Now we can assume without loss of generality that $\alpha(A) > 0$ by selecting ν normal to ∂H.

Then
$$\frac{\beta(A)}{\alpha(A)} = \frac{\nu(A)}{\mu(A)},$$
so it suffices to show that the left hand side is a representation for $\succsim$. Take $A \succsim B$. By construction of ν_A, ν_B,
$$\nu_B(A) \geq 0 = \nu_A(A).$$
Also, our selection of α and β implies

$$\begin{aligned} \nu(A) - \frac{\beta(A)}{\alpha(A)}\mu(A) &= \frac{1}{\alpha(A)}\left[\alpha(A)\nu(A) - \beta(A)\mu(A)\right] \\ &= \frac{\nu_B(A)}{\alpha(A)} \\ &\geq 0, \end{aligned}$$

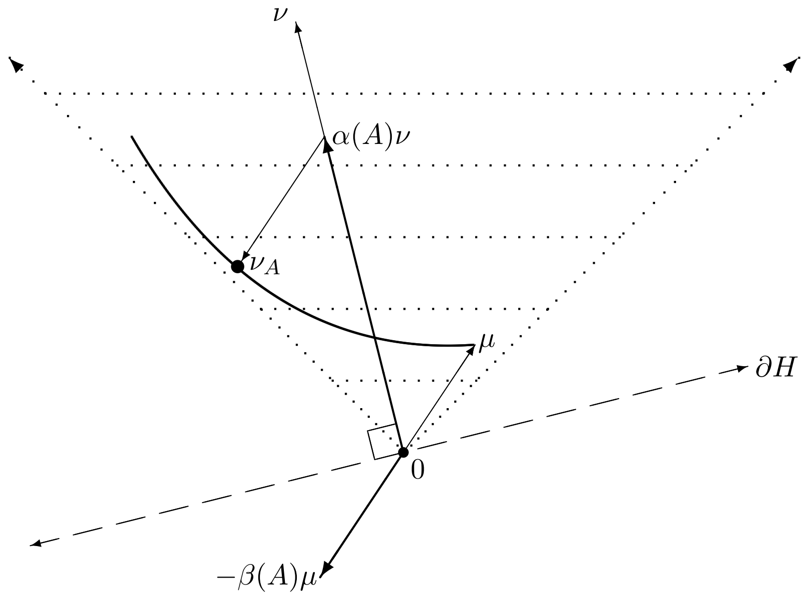

Figure 11: Constructions in the proof of Lemma 29

This two-dimensional figure shows the span of ν and μ, laying the plane in Figure 10 flat against the page. The line ∂H is the border that defines the half space H in which all of the ν_A's live; ν is normal to that border. The coefficients $\alpha(A)$ and $\beta(A)$ on ν and μ are shown for a particular set A; the coefficients are defined by $\nu_A = \alpha(A)\mu - \beta(A)\nu$.

the last inequality following from $\alpha(A) > 0$. Similarly,

$$\nu(A) - \frac{\beta(B)}{\alpha(B)}\mu(A) \leq 0.$$

Together these two inequalities imply

$$\frac{\beta(A)}{\alpha(A)} \geq \frac{\beta(B)}{\alpha(B)}.$$

The argument when $A \succ B$ is identical, replacing weak inequalities with strict inequalities. □

Invoking the Radon–Nikodym Theorem, we can rewrite $\nu(A)$ of Lemma 29 as $\int_A u\, d\mu$, where u is the Radon–Nikodym derivative of ν with respect to μ. The absolute continuity condition on these measures holds from their construction. The continuity of u is guaranteed by the continuity axiom and Lemma 19. The nonatomicity of μ follows from continuity; full support follows from the strict part of disjoint set betweenness. This proves the sufficiency of the axioms for the representation. □

A.4 Proof of Theorem 9

Fix the representation (ϕ, u, μ) delivered by the linearity axiom. The space of probability measures spans the space of all signed measures. So there is a unique extension of $\phi \circ u$ to the space of all signed measures since ϕ is linear. We work with this extension and denote $U(y) = \phi(u(y))$. Furthermore, f refers to the Radon–Nikodym derivative $\frac{d\mu}{d\lambda}$ where λ is the Haar probability measure on ΔX. We first show this derivative exists.

Lemma 30. *The Radon–Nikodym derivative $f = \frac{d\mu}{d\lambda}$ exists.*

Proof. Lemma 20 of Appendix A.3 shows that the extended preference is continuous with respect to the π metric generated by λ on the measure σ-algebra $\mathcal{B}(\Delta X)/\mathcal{N}$. Then μ must be absolutely continuous with respect to λ. If not, there exists a set A with $\lambda(A) > 0$ and a disjoint set $B \succ A$ such that $A \cup B \prec B$. But $\lambda([A \cup B]\Delta B) = 0$, and this contradicts continuity of the extended preference. □

Lemma 31.

$$U(\alpha A + (1-\alpha)x) = \alpha U(A) + (1-\alpha)U(\{x\})$$

and

$$U(A + y) = U(A) + U(y).$$

Proof. The ambiguity-free equivalent of a set A is identified as any point x with $U(\{x\}) = U(A)$. Since ϕ is strictly increasing and linear, it has an inverse, so we can equivalently write $u(x) = \phi^{-1}(U(\{x\}))$. Singleton-set independence implies

$$\phi^{-1}[U(\alpha A + (1-\alpha)x)] = \alpha\phi^{-1}[U(A)] + (1-\alpha)\phi^{-1}[U(\{x\})].$$

Applying ϕ to both sides of the equation yields

$$\begin{aligned} U(\alpha A + (1-\alpha)x) &= \phi\left(\alpha\phi^{-1}[U(A)] + (1-\alpha)\phi^{-1}[U(\{x\})]\right) \\ &= \alpha U(A) + (1-\alpha)U(\{x\}), \end{aligned}$$

where the last step follows from linearity, which allows us to distribute ϕ across the mixture. Recall singleton-set independence implies translation independence by Theorem 7, so

$$\phi^{-1}U(A + y) = \phi^{-1}U(A) + \phi^{-1}U(y).$$

Similarly, this implies $U(A + y) = U(A) + U(y)$. □

Lemma 32.

$$\frac{\mu(A)}{\mu(A+y)} = \frac{\mu(B)}{\mu(B+y)}.$$

Proof. Take any ambiguous sets A and B with $A \succ B$ and any signed measure y with $U(y) \neq 0$. We can assume they are disjoint without loss of generality, by splitting their union into $A \setminus B$, $B \setminus A$, and $A \cap B$. Then

$$U(A \cup B) = \frac{\mu(A)}{\mu(A \cup B)}U(A) + \frac{\mu(B)}{\mu(A \cup B)}U(B).$$

By the second half of Lemma 31:

$$\begin{aligned} U(A \cup B) + U(y) &= U([A \cup B] + y) \\ &= U([A + y] \cup [B + y]) \\ &= \frac{\mu(A+y)}{\mu([A \cup B] + y)}U(A+y) + \frac{\mu(B+y)}{\mu([A \cup B] + y)}U(B+y) \\ &= \left[\frac{\mu(A+y)}{\mu([A \cup B] + y)}U(A) + \frac{\mu(B+y)}{\mu([A \cup B] + y)}U(B)\right] + U(y). \end{aligned}$$

Then

$$U(A \cup B) = \frac{\mu(A+y)}{\mu([A \cup B] + y)}U(A) + \frac{\mu(B+y)}{\mu([A \cup B] + y)}U(B).$$

These two equations imply that the weighting on $U(A)$ in both convex combinations must be equal:

$$\frac{\mu(A)}{\mu(A \cup B)} = \frac{\mu(A+y)}{\mu([A \cup B]+y)}.$$

Rearranging terms,

$$\frac{\mu(A)}{\mu(A+y)} = \frac{\mu(A \cup B)}{\mu([A \cup B]+y)}.$$

Symmetrically,

$$\frac{\mu(B)}{\mu(B+y)} = \frac{\mu(A \cup B)}{\mu([A \cup B]+y)}.$$

Therefore,

$$\frac{\mu(A)}{\mu(A \cup B)} = \frac{\mu(B)}{\mu(B+y)}.$$

□

Lemma 33.

$$\frac{\mu(A)}{\mu(\alpha A+(1-\alpha)x)} = \frac{\mu(B)}{\mu(\alpha B+(1-\alpha)x)}.$$

Proof. Again take A, B disjoint. The first half of Lemma 31 implies:

$$\begin{aligned}
& \alpha U(A \cup B) + (1-\alpha)U(x) \\
= \; & U(\alpha[A \cup B] + (1-\alpha)x) \\
= \; & U([\alpha A + (1-\alpha)x] \cup [\alpha B + (1-\alpha)x]) \\
= \; & \frac{\mu(\alpha A+(1-\alpha)x)}{\mu(\alpha[A \cup B]+(1-\alpha)x)} U(\alpha A + (1-\alpha)x) + \frac{\mu(\alpha B+(1-\alpha)x)}{\mu(\alpha[A \cup B]+(1-\alpha)x)} U(\alpha B + (1-\alpha)x) \\
= \; & \alpha \left[\frac{\mu(\alpha A+(1-\alpha)x)}{\mu(\alpha[A \cup B]+(1-\alpha)x)} U(A) + \frac{\mu(\alpha B+(1-\alpha)x)}{\mu(\alpha[A \cup B]+(1-\alpha)x)} U(B) \right] + (1-\alpha)U(x).
\end{aligned}$$

Then an argument similar to the end of the proof of Lemma 32 delivers the result. □

Lemma 34. *If f is continuous at any point a, it is continuous everywhere.*

Proof. Suppose f is continuous at a. Pick a sequence $A_n \to \{a\}$. By Lemma 19 in Appendix A.3, $\frac{\mu(A_n)}{\lambda(A_n)} \to f(a)$. Any $b \in \Delta X$ can be expressed as the translation $x + y_b$ for some signed measure y_b. Letting $B_n = A_n + y_b$, we show $\frac{\mu(B_n)}{\lambda(B_n)}$ converges for any b. The numerator is a fixed multiple of $\mu(A_n)$ because the ratio $\frac{\mu(B_n)}{\mu(A_n)} = \frac{\mu(A_n+y_b)}{\mu(A_n)}$ is constant by Lemma 32; denote this constant by β. The denominator $\lambda(B_n) = \lambda(A_n + y) = \lambda(A_n)$, by the construction of the Haar measure on ΔX. So $\frac{\mu(B_n)}{\lambda(B_n)} \to \beta\left(\lim \frac{\mu(A_n)}{\lambda(A_n)}\right)$. Then, we can consider the function $f(b) = \lim \frac{\mu(B_n)}{\lambda(B_n)}$ and this limit is robust for any sequence $B_n \to \{b\}$. Now take a sequence of measurable partitions Π_n of ΔX such that for any $b \in \Delta X$, there exists some $B_n \in \Pi_n$ with $B_n \to \{b\}$; for example, the partitions defined by finer grids in the ρ metric. Consider the simple functions defined by

$$f_n(x) = \frac{\mu(B_n)}{\lambda(B_n)}$$

for all $x \in B_n \in \Pi_n$. The definition of the Lebesgue integral implies $f' = \lim f_n$ is a version of the Radon–Nikodym derivative. f' is continuous by construction in view of Lemma 19. So, if f is continuous at a we may take without loss of generality that f is continuous everywhere, as there is a version $f' = f$ almost everywhere with f' continuous everywhere. □

Lemma 35. *f is continuous.*

Proof. We begin by demonstrating Lemma 32 implies that almost surely either $f(a) \geq f(a+y)$ or $f(a) \geq f(a-y)$ for all directions y. Suppose not. Then there is a set A and a direction y with $f(a+y) > f(a)$ for $a \in A$ almost surely and a set B with $f(b+y) < f(b)$ for $b \in B$ almost surely with $\mu(A), \mu(B) > 0$. By the definition of Haar measure, $\mu(A+y) = \int_{A+y} f(x)\, d\lambda > \int_A f(x+y)\, d\lambda = \mu(A)$. Similarly, $\mu(B+y) < \mu(B)$. Then

$$\frac{\mu(A)}{\mu(A+y)} < 1 < \frac{\mu(B)}{\mu(B+y)},$$

which contradicts Lemma 32.

This shows that, almost surely, f is monotone in all directions y of arbitrary length. Then, f is Fréchet differentiable almost everywhere, since all its directional Fréchet differentials exist almost everywhere. It is then continuous almost everywhere (Luenberger 1969, Proposition 3, p. 173). Since this implies there exists at least one point where f is continuous, Lemma 34 implies f is continuous everywhere.[40] □

Necessity is omitted. Since f is continuous by Lemma 35, we can invoke Lemma 19 from Appendix A.3. It implies $f(a) = \lim \frac{\mu(A_n)}{\lambda(A_n)}$. Taking the limit of a sequence $A_n \to \{a\}$ and using Lemma 33:

$$\frac{f(a)}{f(\alpha a + (1-\alpha)x)} = \frac{f(b)}{f(\alpha b + (1-\alpha)x)}.$$

Lemma 32 similarly implies

$$\frac{f(a)}{f(a+y)} = \frac{f(b)}{f(b+y)}.$$

Translating both a and $a+y$ by y yields

$$\frac{f(a)}{f(a+y)} = \frac{f(a+y)}{f(a+2y)}.$$

Taking the convex combinations of $\frac{1}{2}a + \frac{1}{2}(a+2y) = a+y$ and $\frac{1}{2}(a+y) + \frac{1}{2}(a+2y) = a + \frac{3}{2}y$ yields

$$\frac{f(a)}{f(a+y)} = \frac{f(a+y)}{f(a+\frac{3}{2}y)}.$$

Translating $a+y$ and $a+\frac{3}{2}y$ by $\frac{1}{2}y$ yields

$$\frac{f(a+y)}{f(a+\frac{3}{2}y)} = \frac{f(a+\frac{3}{2}y)}{f(a+2y)}.$$

[40] A more elegant but advanced proof of Lemma 35 uses Lusin's Theorem. We include this proof here because it is more direct and independently provides the stronger result that f is almost everywhere Fréchet differentiable.

Combining these equalities:

$$\begin{aligned}\frac{f(a)}{f(a+y)} &= \frac{f(a+y)}{f(a+2y)} \\ &= \frac{f(a+y)}{f(a+\frac{3}{2}y)} \cdot \frac{f(a+\frac{3}{2}y)}{f(a+2y)} \\ &= \left[\frac{f(a+y)}{f(a+\frac{3}{2}y)}\right]^2 \\ &= \left[\frac{f(a)}{f(a+y)}\right]^2\end{aligned}$$

This forces $f(a) = f(a+y)$. Since this is for arbitrary a, y and any point b can be expressed as $a+y$ for some y, we conclude $f(a) = f(b)$ for all a, b. Then the Radon–Nikodym derivative $\frac{d\mu}{d\lambda} = f$ is constant, so μ differs from λ only by a constant scaling. As μ and λ are both probability measures, this suffices to show $\mu = \lambda$. □

A.5 Proof of Propositions 12, 13, and 14

Recall the minimum variance portfolio program:

$$\min \frac{1}{2}\mathbf{w}^\top \mathbf{V}\mathbf{w}$$

subject to

$$\begin{aligned}\mathbf{w}^\top \mathbf{e} &= \mu, \\ \mathbf{w}^\top \mathbf{d} &= \nu, \\ \mathbf{w}^\top \mathbf{1} &= 1.\end{aligned}$$

This is solved by the Lagrangean

$$\min_{\mathbf{w},\lambda_1,\lambda_2,\gamma} L = \frac{1}{2}\mathbf{w}^\top \mathbf{V}\mathbf{w} + \lambda_1(\mu - \mathbf{w}^\top \mathbf{e}) + \lambda_2(\nu - \mathbf{w}^\top \mathbf{d}) + \gamma(1 - \mathbf{w}^\top \mathbf{1}).$$

The first order conditions are:

$$\begin{aligned}\frac{\partial L}{\partial \mathbf{w}} &= \mathbf{V}\mathbf{w} - \gamma\mathbf{1} - \lambda_1\mathbf{e} - \lambda_2\mathbf{d} = 0 \\ \frac{\partial L}{\partial \lambda_1} &= \mu - \mathbf{w}^\top \mathbf{e} = 0 \\ \frac{\partial L}{\partial \lambda_2} &= \nu - \mathbf{w}^\top \mathbf{d} = 0 \\ \frac{\partial L}{\partial \gamma} &= 1 - \mathbf{w}^\top \mathbf{1} = 0\end{aligned}$$

These imply:

$$\mathbf{w} = \lambda_1 \mathbf{V}^{-1}\mathbf{e} + \lambda_2 \mathbf{V}^{-1}\mathbf{d} + \gamma \mathbf{V}^{-1}\mathbf{1}. \quad (1)$$

Premultiplying this equation by $\mathbf{e}, \mathbf{d}$, and $\mathbf{1}$ yields three new equations, which we write in matrix form:

$$\begin{bmatrix} \mathbf{e}^\top \mathbf{V}^{-1}\mathbf{e} & \mathbf{d}^\top \mathbf{V}^{-1}\mathbf{e} & \mathbf{1}^\top \mathbf{V}^{-1}\mathbf{e} \\ \mathbf{e}^\top \mathbf{V}^{-1}\mathbf{d} & \mathbf{d}^\top \mathbf{V}^{-1}\mathbf{d} & \mathbf{1}^\top \mathbf{V}^{-1}\mathbf{d} \\ \mathbf{e}^\top \mathbf{V}^{-1}\mathbf{1} & \mathbf{d}^\top \mathbf{V}^{-1}\mathbf{1} & \mathbf{1}^\top \mathbf{V}^{-1}\mathbf{1} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \gamma \end{bmatrix} = \begin{bmatrix} \mu \\ \nu \\ 1 \end{bmatrix} \tag{2}$$

Premultiplying by the inverse of the square matrix in (2) yields

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \gamma \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} D'F' - E'^2 & C'E' - B'F' & B'E' - C'D' \\ C'E' - B'F' & A'F' - C'^2 & B'C' - A'E' \\ B'E' - C'D' & B'C' - A'E' & A'D' - B'^2 \end{bmatrix} \begin{bmatrix} \mu \\ \nu \\ 1 \end{bmatrix}, \tag{3}$$

where

$$\begin{aligned} A' &= \mathbf{e}^\top \mathbf{V}^{-1}\mathbf{e} \\ B' &= \mathbf{d}^\top \mathbf{V}^{-1}\mathbf{e} = \mathbf{e}^\top \mathbf{V}^{-1}\mathbf{d} \\ C' &= \mathbf{1}^\top \mathbf{V}^{-1}\mathbf{e} = \mathbf{e}^\top \mathbf{V}^{-1}\mathbf{1} \\ D' &= \mathbf{d}^\top \mathbf{V}^{-1}\mathbf{d} \\ E' &= \mathbf{1}^\top \mathbf{V}^{-1}\mathbf{d} = \mathbf{d}^\top \mathbf{V}^{-1}\mathbf{1} \\ F' &= \mathbf{1}^\top \mathbf{V}^{-1}\mathbf{1} \\ \Delta &= A'D'F' + 2B'C'E' - A'E'^2 - D'C'^2 - F'B'^2. \end{aligned}$$

We can rewrite the right hand side of (3) and solve for the multipliers:

$$\begin{aligned} \lambda_1 &= \frac{A\mu + B\nu + C}{\Delta} \\ \lambda_2 &= \frac{B\mu + D\nu + E}{\Delta} \\ \gamma &= \frac{C\mu + E\nu + F}{\Delta}, \end{aligned}$$

letting

$$\begin{aligned} A &= D'F' - E'^2 \\ B &= C'E' - B'F' \\ C &= B'E' - C'D' \\ D &= A'F' - C'^2 \\ E &= B'C' - A'E' \\ F &= A'D' - B'^2. \end{aligned}$$

Let

$$\begin{aligned}
\mathbf{f} &= \frac{C\mathbf{V}^{-1}\mathbf{e} + E\mathbf{V}^{-1}\mathbf{d} + F\mathbf{V}^{-1}\mathbf{1}}{\Delta} \\
\mathbf{g} &= \frac{A\mathbf{V}^{-1}\mathbf{e} + B\mathbf{V}^{-1}\mathbf{d} + C\mathbf{V}^{-1}\mathbf{1}}{\Delta} \\
\mathbf{h} &= \frac{B\mathbf{V}^{-1}\mathbf{e} + D\mathbf{V}^{-1}\mathbf{d} + E\mathbf{V}^{-1}\mathbf{1}}{\Delta}.
\end{aligned}$$

Substituting our solutions for the multipliers and these abbreviations for (1) yields

$$\mathbf{w}^*(\mu, \nu) = \mathbf{f} + \mathbf{g}\mu + \mathbf{h}\nu.$$

This proves Proposition 12: the optimal weights are linear in μ and ν.

But $\mathbf{f}, \mathbf{f}+\mathbf{g}$, and $\mathbf{f}+\mathbf{h}$ are all frontier portfolios, generated by (μ, ν) values of $(0,0), (1,0)$, and $(0,1)$:

$$\begin{aligned}
\mathbf{f} + \mathbf{g}\cdot 0 + \mathbf{h}\cdot 0 &= \mathbf{f} \\
\mathbf{f} + \mathbf{g}\cdot 1 + \mathbf{h}\cdot 0 &= \mathbf{f} + \mathbf{g} \\
\mathbf{f} + \mathbf{g}\cdot 0 + \mathbf{h}\cdot 1 &= \mathbf{f} + \mathbf{h}
\end{aligned}$$

Then, by placing weights of $(1-\mu-\nu, \mu, \nu)$ on these portfolios, we can generate any portfolio on the frontier. This proves Proposition 14, since these three portfolios are the spanning base for the frontier.

We now solve for variance generated by the minimum variance portfolio as a function of μ and ν.

$$\begin{aligned}
\sigma^2 &= \mathbf{w}^\top \mathbf{V}\mathbf{w} \\
&= \mathbf{w}^\top \mathbf{V}(\lambda_1 \mathbf{V}^{-1}\mathbf{e} + \lambda_2 \mathbf{V}^{-1}\mathbf{d}\mathbf{V}^{-\mathbf{1}}\mathbf{1}) \\
&= \lambda_1 \mathbf{w}^\top \mathbf{e} + \lambda_2 \mathbf{w}^\top \mathbf{d} + \gamma_2 \mathbf{w}^\top \mathbf{1} \\
&= \lambda_1 \mu + \lambda_2 \nu + \gamma \\
&= \frac{1}{\Delta}\left([A\mu + B\nu + C]\,\mu + [B\mu + D\nu + E]\,\nu + [C\mu + E\nu + F]\right) \\
&= \frac{A\mu^2 + D\nu^2 + 2B\mu\nu + 2C\mu + 2E\nu + F}{\Delta}
\end{aligned}$$

The mean-variance-ambiguity frontier is thus a paraboloid in $\mathbb{R}^3$, proving Proposition 13.

$$\begin{aligned}
\frac{\partial \sigma^2}{\partial \mu} &= \frac{2(A\mu + B\nu + C)}{\Delta} \\
\frac{\partial \sigma^2}{\partial \nu} &= \frac{2(B\mu + D\nu + E)}{\Delta}
\end{aligned}$$

Then the minimum variance portfolio is attained at $\mu = \frac{BE-CD}{AD-B^2}$ and $\nu = \frac{BC-EA}{AD-B^2}$.

$$\begin{aligned}
\mathrm{Cov}(r_p, r_q) &= \mathbf{w}_p^\top \mathbf{V}\mathbf{w}_q \\
&= \lambda_1 \mathbf{e}^\top \mathbf{V}^{-1}\mathbf{V}\mathbf{w}_q + \lambda_2 \mathbf{d}^\top \mathbf{V}^{-1}\mathbf{V}\mathbf{w}_q + \gamma \mathbf{1}^\top \mathbf{V}^{-1}\mathbf{V}\mathbf{w}_q \\
&= \lambda_1 \mu_q + \lambda_2 \nu_q + \gamma \\
&= \left(\frac{A\mu_p + B\nu_p + C}{\Delta}\right)\mu_q + \left(\frac{B\mu_p + D\nu_p + E}{\Delta}\right)\nu_q + \frac{C\mu_p + E\nu_p + F}{\Delta}
\end{aligned}$$

References

Ahn, D. S. (2003): "Hierarchies of Ambiguous Beliefs," Working paper, Stanford Graduate School of Business.

Aliprantis, C. D., and K. C. Border (1999): *Infinite Dimensional Analysis*. Springer, New York, 2nd edn.

Allais, M. (1953): "Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Américaine," *Econometrica*, 21, 503–546.

Anscombe, F. J., and R. J. Aumann (1963): "A Definition of Subjective Probability," *Annals of Mathematical Statistics*, 34, 199–205.

Arrow, K. J., and L. Hurwicz (1972): "An Optimality Criterion for Decision-Making under Ignorance," in *Uncertainty and Expectations in Economics: Essays in Honour of G. L. S. Shackle*, ed. by C. F. Carter, and J. L. Ford, pp. 1–11. Blackwell, Oxford.

Birkhoff, G., and J. von Neumann (1936): "The Logic of Quantum Mechanics," *Annals of Mathematics*, 37, 823–843.

Blume, L., A. Brandenburger, and E. Dekel (1991): "Lexicographic Probabilities and Choice Under Uncertainty," *Econometrica*, 59, 61–79.

Bolker, E. D. (1966): "Functions Resembling Quotients of Measures," *Transactions of the American Mathematial Society*, 124, 292–312.

——— (1967): "A Simultaneous Axiomatization of Utility and Subjective Probability," *Philosophy of Science*, 34, 333–340.

Broome, J. (1990): "Bolker–Jefrey Expected Utility Theory and Axiomatic Utilitarianism," *Review of Economic Studies*, 57, 477–502.

Casadesus-Masanell, R., P. Klibanoff, and E. Ozdenoren (2000): "Maxmin Expected Utility over Savage Acts with a Set of Priors," *Journal of Economic Theory*, 92, 35–65.

Damiano, E. (1999): "Choice under Limited Uncertainty," Working paper, Yale University.

Debreu, G. (1954): "Representation of a Preference Ordering by a Numerical Function," in *Decision Processes*, ed. by R. M. Thrall, C. H. Coombs, and R. L. Davis. Wiley, New York.

Dekel, E., B. L. Lipman, and A. Rustichini (2001): "Representing Preferences with a Unique Subjective State Space," *Econometrica*, 69, 891–934.

Doherty, N. A., and S. M. Tinic (1981): "Reinsurance under Conditions of Capital Market Equilibrium: A Note," *Journal of Finance*, 36, 949–953.

Ellsberg, D. (1961): "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics*, 75, 643–669.

Epstein, L. G. (1999): "A Definition of Uncertainty Aversion," *Review of Economic Studies*, 66, 579–608.

Epstein, L. G., and J. Zhang (2001): "Subjective Probabilities on Subjectively Unambiguous Events," *Econometrica*, 69(2), 265–306.

Ergin, H., and F. Gul (2002): "A Subjective Theory of Compound Lotteries," Working paper, Princeton University.

Gajdos, T., J.-M. Tallon, and J.-C. Vergnaud (2003): "Decision Making with Imprecise Probabilistic Information," Working paper, CNRS–Universitè Paris I Panthèon–Sorbonne.

Ghirardato, P., F. Maccheroni, and M. Marinacci (2002): "Ambiguity from the Differential Viewpoint," Social Science Working Paper #1130, California Institute of Technology.

Ghirardato, P., and M. Marinacci (2002): "Ambiguity Made Precise: A Comparative Foundation," *Journal of Economic Theory*, 102, 251–289.

Gilboa, I., and D. Schmeidler (1989): "Maxmin Expected Utility with a Non–Unique Prior," *Journal of Mathematical Economics*, 18, 141–153.

——— (1995): "Case–Based Decision Theory," *Quarterly Journal of Economics*, 110, 605–639.

Gul, F., and W. Pesendorfer (2001): "Temptation and Self–Control," *Econometrica*, 69, 1403–1435.

Halmos, P. R. (1974): *Measure Theory*. Springer, New York, Reprint.

Hayashi, T. (2003): "Information, Subjective Belief and Preference," Working paper, University of Rochester.

Jeffrey, R. C. (1965): *The Logic of Decision*. McGraw-Hill, New York.

Klein, E., and A. C. Thompson (1984): *Theory of Correspondences*. Wiley, New York.

Klibanoff, P. (1996): "Uncertainty, Decision, and Normal Form Games," Working paper, Kellogg Graduate School of Management, Northwestern University.

Klibanoff, P., M. Marinacci, and S. Mukerji (2003): "A Smooth Model of Decision Making under Ambiguity," Working paper, Kellogg Graduate School of Management, Northwestern University.

Kraft, C. H., J. W. Pratt, and A. Seidenberg (1959): "Intuitive Probability on Finite Sets," *Annals of Mathematical Statistics*, 30, 408–419.

Krantz, D. H., R. D. Luce, P. Suppes, and A. Tversky (1971): *Foundations of Measurement*, vol. 1. Academic Press, New York.

Lo, K. C. (1996): "Equilibrium in Beliefs under Uncertainty," *Journal of Economic Theory*, 71, 443–484.

Luenberger, D. G. (1969): *Optimization by Vector Space Methods*. Wiley, New York.

Mayers, D., and C. W. Smith, Jr. (1990): "On the Corporate Demand for Insurance: Evidence from the Reinsurance Market," *Journal of Business*, 63, 19–40.

McCullough, R. C. (1964): "Fables of Reinsurance," *Journal of Risk and Insurance*, 31, 259–263.

Nash, J. F. (1950): "Equilibrium Points in N-Person Games," *Proceedings of the National Academy of Sciences*, 36, 48–49.

Nau, R. F. (2003): "Uncertainty Aversion with Second–Order Utilities and Probabilities," Working paper, Fuqua School of Business, Duke University.

Olszewski, W. (2003): "Preferences over Sets of Lotteries," Working paper, Northwestern University.

Parthasarathy, K. R. (1967): *Probability Measures on Metric Spaces*. Academic Press, New York.

Savage, L. J. (1954): *The Foundations of Statistics*. Wiley, New York.

Schmeidler, D. (1989): "Subjective Probability and Expected Utility without Additivity," *Econometrica*, 57, 571–587.

Siniscalchi, M. (2003): "A Behavioral Characterization of Plausible Priors," Working paper, Northwestern University.

Stinchcombe, M. B. (2003): "Choice and Games with Ambiguity as Sets of Probabilities," Working paper, University of Texas–Austin.

Suppes, P. (1966): "The Probabilistic Argument for a Non-Classical Logic of Quantum Mechanics," *Philosophy of Science*, 33, 14–21.

von Neumann, J. (1955): *Mathematical Foundations of Quantum Mechanics*. Princeton University Press, Princeton, translated by R. T. Beyer.

von Neumann, J., and O. Morgenstern (1944): *Theory of Games and Economic Behavior*. Princeton University Press, Princeton.

Wang, T. (2003): "A Class of Multi–Prior Preferences," Working paper, University of British Columbia.

Zhang, J. (1999): "Qualitative Probabilities on λ-Systems," *Mathematical Social Sciences*, 38, 11–20.