Theoretical Economics 16 (2021), 539–570 1555-7561/20210539

Random ambiguity

Jay Lu
Department of Economics, UCLA

We introduce a model of random ambiguity aversion. Choice is stochastic due to unobserved shocks to both information and ambiguity aversion. This is modeled as a random set of beliefs in the maxmin expected utility model of Gilboa and Schmeidler (1989). We characterize the model and show that the distribution of ambiguity aversion can be uniquely identified from binary choices. A novel stochastic order on random sets is introduced that characterizes greater uncertainty aversion under stochastic choice. If the set of priors is the Aumann expectation of the random set, then choices satisfy dynamic consistency. This corresponds to an agent who knows the distribution of signals but is uncertain about how to interpret signal realizations. More broadly, the analysis of stochastic properties of random ambiguity attitudes provides a theoretical foundation for the study of other random nonlinear utility models.

Keywords. Stochastic choice, ambiguity, random utility, updating.

JEL classification. D81, D83.

1. Introduction

In many economic situations, ambiguity aversion, i.e., aversion toward Knightian uncertainty (Knight 1921), is useful for explaining behavior that cannot be easily addressed by standard expected utility maximization. In practice, however, it is useful to model choice behavior as stochastic due to unobserved heterogeneity in a population of agents or unobserved shocks to a single agent's preferences. In this paper, we provide a theory of stochastic choice generated by random ambiguity.

To be concrete, consider a population of agents choosing between a safe and an uncertain asset where prices are fixed. Agents have heterogeneous ambiguity attitudes.[1] Those who are more ambiguity-averse choose the safe asset while those who are less ambiguity averse choose the uncertain one.[2] An analyst observes only the proportion of

Jay Lu: jay@econ.ucla.edu
This is a significantly revised version of Chapter 2 of my dissertation at Princeton. I am deeply indebted to Faruk Gul and Wolfgang Pesendorfer for their advice, guidance, and encouragement. I would also like to thank Larry Blume, David Easley, Diogo Guillen, Rohit Lamba, Yi-Hsuan Lin, Stephen Morris, Paulo Natenzon, Pietro Ortoleva, and Kota Saito for their helpful comments. Financial support from the NSF under Award SES-1919275 is gratefully acknowledged.

[1]Empirical evidence of this heterogeneity include Abdellaoui et al. (2011) and Ahn et al. (2014).

[2]One could also interpret the magnitude of ambiguity aversion as reflective of uncertainty about the unknown state. For example, in Caballero and Krishnamurthy (2008), agents face an enlarged set of possible priors during surprise defaults or bank runs and scramble for safe assets in a flight-to-quality.

 https://doi.org/10.3982/TE3810

agents who choose each asset (i.e., stochastic choice), but not the distribution of ambiguity aversion. Our model shows how the analyst can uniquely identify this distribution using stochastic choice from a rich set of binary options. This is an important exercise if the analyst is a regulator deciding on optimal policy implementation (see Easley and O'Hara 2009).

Alternatively, consider a single hiring manager faced with a pool of job applicants. The manager conducts private interviews before deciding whether to hire each worker. Hiring decisions depend on the information acquired from interviews as well as the manager's perceived ambiguity regarding each worker. An analyst observes only the hiring rate (i.e., stochastic choice) of the manager, but not the distribution of interview outcomes. Our model shows how the analyst can identify this distribution using stochastic choice.

We introduce a model of stochastic choice and ambiguity that captures both examples. In the multiple-priors model of Gilboa and Schmeidler (1989), an agent evaluates an option f by minimizing expected utility over a set K of possible beliefs about the payoff-relevant state. The maxmin expected utility of the option f is given by

$$u_K(f) = \min_{\pi \in K} \pi \cdot (u \circ f).$$

In our *random ambiguity* model, the von Neumann–Morgenstern (vNM) utility u is fixed, but the set of beliefs K is random. In other words, this is a random utility model where the utilities are maxmin expected utilities that depend on K. The distribution of ambiguity aversion is captured by the distribution of the random set K.

Our main results are as follows. First, we show that the distribution of the random set K can be fully identified from stochastic choice from binary menus. We then introduce a novel stochastic order on random sets that characterizes greater uncertainty aversion in this context. Next, we show that if the ex ante set of priors is exactly the Aumann expectation of K (i.e., the set of all priors consistent with K), then choice behavior satisfies dynamic consistency. We also provide a full axiomatic characterization of the model. Although we focus on the maxmin model due to its tractability, our analysis of stochastic properties of random ambiguity attitudes provides a theoretical foundation for the general study of random nonlinear utility.

Section 2.1 introduces the formal setup. Let S be a finite payoff-relevant state space. An act is a state-contingent mapping from S to payoffs in the form of risky prospects. Following Anscombe and Aumann (1963) and Seo (2009), we distinguish between randomization that is ex ante (i.e., before the state is realized) and ex post (i.e., after the state is realized). Formally, each choice option is a lottery, an ex ante mixture over acts. The main primitive is a stochastic choice over menus of lotteries that specify the choice probability for each lottery p in the menu A. In our model, the probability of choosing a lottery $p \in A$ is precisely the probability that p attains the highest utility in the menu A. Since ambiguity is with respect to the state, ex ante mixtures are evaluated using standard expected utility. Thus, this is a random utility model where utilities are expected utilities with respect to ex ante randomization and maxmin with respect to ex post randomization.

Distinguishing between ex ante and ex post randomization in this stochastic choice model provides several advantages. First, the use of ex ante randomization facilitates the axiomatic characterization (see Section 3) as characterizing random utility in general without additional structure is a known difficulty.[3] Second, the use of lotteries over acts allows for sharper identification results (Section 2.2) and cleaner comparative statics (Section 2.3).

Section 2.2 considers identification. Theorem 1 shows that the distribution of the random set K can be identified from simple binary menus where one option is constant (i.e., yields the same payoff regardless of the state). This is achieved by leveraging the richness of ex ante randomization. In the absence of ex ante randomization, identification can still be achieved, but requires stochastic choice data beyond binary menus (see Theorem 7 in the Appendix); this is possible despite the well known difficulties in identifying models of random nonexpected utility (see Lin 2020).

Section 2.3 considers comparative statics. Call one stochastic choice more uncertainty-averse than another if constant acts are chosen more frequently in the first than in the second. Theorem 2 shows that greater uncertainty aversion is exactly characterized by a novel stochastic order called stochastic c-dominance. A random set of beliefs is greater than another in this order if it puts greater weight on all increasing, convex, and closed families of beliefs. This notion is weaker than the standard stochastic order on random sets that corresponds to first-order stochastic dominance with respect to set inclusion.

In Section 2.4, we extend our model to address updating and dynamic consistency. Consider a random set of beliefs K. If a random belief is always in the random set, then we call it a posterior selection of K. Suppose the agent's initial set of priors is the Aumann expectation of K, i.e., the set of priors of all posterior selections of K. This could correspond to an agent who knows the distribution of signals but is uncertain about how to interpret signal realizations. As a result, he has in mind a family of possible signal structures. Theorem 3 shows that if the agent's ex ante preferences are represented by a maxmin expected utility where the set of ex ante priors is the Aumann expectation, then the agent satisfies dynamic consistency. This indicates that stochastic models of ambiguity may suggest new approaches toward updating with multiple priors.

Section 3 provides the axiomatic characterization. The first axiom (monotonicity, ex ante independence, ex ante extremeness, continuity) consists of conditions that characterize random expected utility from Gul and Pesendorfer (2006). A novel axiom, *ex post hedging*, characterizes stochastic aversion toward uncertainty. It states that adding an act will not affect the choice of another act as long as a hedge option (i.e., mixture of the two acts) is available in the menu. *Certainty reversal of order* states that when mixing with constant acts, ex ante and ex post randomization are the same. This is because constant acts involve no uncertainty, so the agent is indifferent to the timing of randomization. Finally, the last three axioms (certainty determinism, dominance, and nondegeneracy) are conditions from the information representation of Lu (2016). Theorem 4 states that these axioms exactly characterize the random ambiguity model.

[3]For instance, compare the characterization of Falmagne (1978) with that of Gul and Pesendorfer (2006) in the richer lottery setup. The use of ex ante randomization allows us to adopt the latter approach.

Finally, in Section 4, we study the general stochastic properties of nonlinear random utility. In the random ambiguity model, Bernoulli utilities are maxmin expected utility and exhibit ambiguity aversion. As a result, ex post randomization is desirable and utilities are quasiconcave (with respect to ex post mixtures). We show that for random utilities, quasiconcavity is exactly characterized by the ex post hedging axiom. Alternatively, if the agent is ambiguity-loving, then ex post randomization is not desirable and utilities are quasiconvex. For random utility, quasiconvexity is exactly characterized by the ex post version of the extremeness axiom of Gul and Pesendorfer (2006). Theorem 5 thus provides a characterization of quasiconcavity and quasiconvexity for random utilities. We then apply the result to characterize betweenness for random utility (Corollary 1) and ambiguity neutrality in our random ambiguity model (Corollary 2), where choice is stochastic only due to information. More generally, these results provide a theoretical foundation for studying other stochastic ambiguity models beyond random maxmin expected utility.

All proofs are provided in the Appendix unless stated otherwise.

1.1 *Related literature*

This paper is related to a long literature on stochastic choice and random utility. Early seminal works include Block and Marschak (1960), McFadden and Richter (1990), and Falmagne (1978). More recent works have leveraged enriched choice domains to obtain more structured characterizations. These include Gul and Pesendorfer (2006), Ahn and Sarver (2013), Fudenberg and Strzalecki (2015), Lu (2016), Lu and Saito (2018), Duraj (2018), Frick et al. (2019), and Lin (2019).

In particular, this paper contributes to our understanding of random nonlinear utility. Violations of linearity in a stochastic context have been well documented (see Becker et al. 1963 and Kahneman and Tversky 1979). In Section 4, we study the stochastic properties of random nonlinear utility and provide characterizations of random quasiconcave and quasiconvex utility. Recently, Lin (2020) shows that contrary to random linear utility, models of random nonlinear utility may lack unique identification.

Recent papers that have also studied ambiguity and stochastic choice include Fudenberg et al. (2015), Saito (2015), and Agranov and Ortoleva (2017). In all these models, the agent deliberately randomizes over choice options due to a preference for ex ante randomization. In contrast, the agent in our model has a preference for ex post randomization and any observed stochastic choice is due to random ambiguity. For stochastic binary choices, Ryan (2018) characterizes a Fechner model of ambiguity aversion.

Epstein and Kopylov (2007) study a model of menu choice where a sophisticated agent anticipates future shocks to ambiguity aversion. While their model is ex ante where choices are due to the anticipation of cold feet shocks, our model is ex post where choices are the direct result of such shocks. Seo (2009) considers a deterministic choice model that also distinguishes between ex ante and ex post randomization. He also relaxes the reversal of order axiom and assumes the agent is impartial to ex ante randomization.

The paper is also related to a large literature on updating under ambiguity. Approaches include Gilboa and Schmeidler (1993), Sarin and Wakker (1998), Epstein and

Schneider (2003), Hanany and Klibanoff (2007), and Siniscalchi (2011). When updating with multiple priors, a tension arises between accommodating rich ambiguity attitudes and satisfying dynamic consistency and consequentialism. We show that if the original set of priors is the Aumann expectation of the random set of posteriors, then the agent satisfies dynamic consistency and consequentialism. Thus, stochastic choice models with random sets of beliefs may suggest new restrictions on updating under ambiguity.

Finally, this paper is related to a large literature on applications of heterogeneous ambiguity aversion. Dow and da Costa Werlang (1992) and Epstein and Wang (1994) study the effects of ambiguity aversion on market nonparticipation and asset prices. Bose et al. (2006) investigate varying ambiguity aversion in an auction setting. Caballero and Krishnamurthy (2008) study investors who receive random shocks to ambiguity aversion due to surprise defaults or bank runs and scramble to safer assets in a phenomenon called flight-to-quality. Easley and O'Hara (2009) study the role of regulation in a heterogeneous population with different levels of ambiguity aversion. Epstein and Schneider (2010) provide a review of various limited participation problems involving heterogeneous ambiguity aversion.

2. A model of random ambiguity

2.1 *Setup and model*

Let S be a finite state space and let X be a finite set of prizes. Let ΔS and ΔX be their respective probability simplexes. We interpret ΔS as the set of all beliefs about S and interpret ΔX as the set of payoffs that consists of risky prospects. As in Anscombe and Aumann (1963), an *act* corresponds to a state-contingent payoff $f : S \to \Delta X$. An act is *constant* if $f(s)$ is the same for all $s \in S$. Let H denote the set of all acts and let $H_c \subset H$ denote the set of constant acts.

Following Anscombe and Aumann (1963) and Seo (2009), we consider lotteries (i.e., Borel probability measures) over acts. Let ΔH denote the set of all lotteries. Given an act $f \in H$, let $\delta_f \in \Delta H$ denote the degenerate lottery that yields f for sure. We distinguish between ex ante and ex post randomization. For instance, given two lotteries $\delta_f, \delta_g \in \Delta H$ and $a \in [0, 1]$, the mixture

$$p = a\delta_f \oplus (1 - a)\delta_g \in \Delta H$$

corresponds to ex ante mixing of payoffs where p yields the act f with probability a and yields g with probability $1 - a$. Since randomization occurs before the state is realized, it is ex ante. Alternatively, given two acts $f, g \in H$ and $a \in [0, 1]$, the mixture

$$h = af + (1 - a)g \in H$$

corresponds to ex post mixing of payoffs where $h(s) = af(s) + (1 - a)g(s)$ for all states $s \in S$. Here, randomization is ex post as it occurs after the state is realized.

Call a finite set of lotteries a *menu* and let $\mathcal{A}$ denote the set of all menus. We endow ΔH with the topology of weak convergence and endow $\mathcal{A}$ with the corresponding Hausdorff metric. Note that $\mathcal{A}$ is a mixture space under Minkowski mixing for ex ante randomization.[4] Stochastic choice corresponds to a choice distribution across all lotteries in a menu. Let $\Delta(\Delta H)$ denote the set of all choice distributions over lotteries.

Definition 1. A *stochastic choice* is a mapping $\rho : \mathcal{A} \to \Delta(\Delta H)$ such that $\rho_A(A) = 1$ for all $A \in \mathcal{A}$.

For any menu $A \in \mathcal{A}$ and lottery $p \in A$, $\rho_A(p)$ specifies the probability that the lottery p is chosen in the menu A. For binary menus $A = \{p, q\}$, we employ the more succinct notation $\rho(p, q) = \rho_A(p)$.

As with any model of random utility, there is a need to address ties. Following Lu (2016), we model ties by relaxing the restriction that all choice probabilities have to be fully specified. For example, if two lotteries p and q are tied, then the stochastic choice does not specify choice probabilities for either p or q specifically. Formally, we model this as nonmeasurability with respect to the choice probability and let ρ denote the corresponding outer measure without loss of generality.[5] With this definition, $\rho(p, q) = \rho(q, p) = 1$ whenever p and q are tied.

In the classic Raiffa (1961) critique, an agent who is ambiguity-averse may strictly prefer randomization for hedging purposes. Differentiating between ex ante and ex post randomization is one way the literature has resolved this issue (see Saito 2015). Since ambiguity is with respect to the state, ex ante randomization that occurs before the state is realized does not help. Stochastic choice adds a third layer of randomization, as the agent can now deliberately randomize between acts (see Cerreia-Vioglio et al. 2019). However, since deliberate randomization also occurs before the state is realized, an agent who is impartial to ex ante randomization would likely also be impartial to deliberate randomization. Adopting this approach allows us to interpret any observed stochastic choice as the result of random ambiguity rather than deliberate randomization.

We now describe the random ambiguity model. Let $\mathcal{K}$ be the set of all nonempty compact convex subsets of ΔS endowed with the Hausdorff metric. Note that $\mathcal{K}$ is also a mixture space under Minkowski mixing. In the multiple-priors model of Gilboa and Schmeidler (1989), maxmin expected utility is characterized by a von Neumann–Morgenstern (vNM) utility $u : \Delta X \to \mathbb{R}$ and a set of beliefs $K \in \mathcal{K}$. Formally, the expected utility of an act $f \in H$ given a belief $\pi \in \Delta S$ is

$$\pi \cdot (u \circ f) = \sum_s \pi(s) u\big(f(s)\big).$$

[4]The Minkowski mixture of two menus $A, B \in \mathcal{A}$ and $a \in [0, 1]$ is given by $aA \oplus (1-a)B = \{ap \oplus (1-a)q : p \in A, q \in B\}$.

[5]Stochastic choice naturally subsumes a σ-algebra $\mathcal{B}$ on ΔH. Given any menu $A \in \mathcal{A}$, the corresponding choice distribution ρ_A is a measure on the σ-algebra generated by $\mathcal{B} \cup \{A\}$. We let ρ denote the outer measure with respect to this σ-algebra without loss. For details, see Lu (2016).

Maxmin expected utility evaluates each act according to the worst possible belief in the set $K \in \mathcal{K}$. In other words, the maxmin expected utility of an act f is given by

$$u_K(f) := \min_{\pi \in K} \pi \cdot (u \circ f).$$

We extend the utility to lotteries by defining the expected utility $U_K(p) := \int_H u_K(f)\,dp$, where the Bernoulli utility $u_K(\cdot)$ is maxmin. Note that this implies that the agent is indifferent to ex ante randomization as in Seo (2009).

In our model, choice is stochastic due to unobserved shocks to both information and ambiguity attitudes. We model this by allowing the set of beliefs K to be random. Formally, let μ be a Borel probability measure on $\mathcal{K}$ with the σ-algebra induced by the Hausdorff metric. A random element taking values in $\mathcal{K}$ with distribution μ is called a *random set*.

Given a vNM utility u, a distribution μ is *regular* if the utilities of two options are either always or never equal, i.e., $U_K(p) = U_K(q)$ with probability 0 or 1. This relaxes the standard restriction in standard random utility models that assumes that utilities are never equal. Going forward, let (μ, u) consist of a regular distribution μ and a nonconstant u. Define a random ambiguity representation as follows.

Definition 2 (Random ambiguity). Stochastic choice ρ is represented by (μ, u) if

$$\rho_A(p) = \mu\big(\{K \in \mathcal{K} : U_K(p) \geq U_K(q) \text{ for all } q \in A\}\big).$$

This is a random utility model where the random Bernoulli utilities are maxmin expected utilities that depend on the random set of beliefs K. The probability of choosing $p \in A$ is precisely the probability that p attains the highest utility in the menu A.

In the group interpretation of the model, there is a population of maxmin agents, and stochastic choice reflects unobserved heterogeneity in both information and ambiguity in the population. This heterogeneity is captured by μ, which is the population distribution of sets of priors.[6] For example, suppose f corresponds to a safe asset (e.g., a constant act), while g corresponds to an asset with uncertain payoffs. In this case, $\rho(\delta_f, \delta_g)$ reflects the proportion of agents in the population who are sufficiently ambiguity-averse (i.e., large enough sets of priors) that they choose the safe asset over the uncertain asset.

In the individual interpretation of the model, there is a single maxmin agent, and stochastic choice reflects the agent's private information and ambiguity attitudes. This is captured by μ, which is a distribution of signals where each signal realization corresponds to a set of priors for the agent. Choice frequencies can be calculated from repeated choices over a series of independent decisions or over time. For example, suppose the agent is a hiring manager, and the acts f and g correspond, respectively, to

[6]Note that in the maxmin model, the size of the set of priors can be interpreted either as the magnitude of ambiguity aversion or the amount of perceived subjective uncertainty about the state. Ghirardato et al. (2004) show that the two interpretations can be distinguished with additional restrictions on choice data. A similar exercise can be performed here.

hiring and not hiring a worker from a pool of job applicants. In this case, $\rho(\delta_f, \delta_g)$ reflects the hiring rate of the manager, which depends on ambiguity attitudes as well as private information gleaned from interviews. Alternatively, suppose f corresponds to constant consumption every period, while g corresponds to uncertain consumption every period. In this case, $\rho(\delta_f, \delta_g)$ reflects the time-averaged long-run frequency that the agent is sufficiently ambiguity-averse and chooses constant consumption (see Lu and Saito 2020).

The following examples are two special cases of the random ambiguity model. In the first example, information is fixed but ambiguity attitudes are random, while in the second, ambiguity attitudes are fixed but information is random.

Example 1. Suppose the random set of beliefs satisfies

$$K = (1-\epsilon)\{\pi\} + \epsilon J,$$

where $\epsilon \in [0, 1]$ is a random parameter, $\pi \in \Delta S$ is a fixed belief, and $J \in \mathcal{K}$ is a fixed set of beliefs such that $\pi \in J$. This is a simple one-dimensional parametrization of random ambiguity aversion where ε measures the dispersion of beliefs from the fixed belief π. It is a stochastic version of the classic ϵ-contamination model that has been widely used in robust statistics.[7] Note that given this parametrization of K, it is straightforward to rewrite the maxmin expected utility as

$$u_K = u_{\{\pi\}} - \epsilon(u_{\{\pi\}} - u_J).$$

The random utility can be thus decomposed into a fixed subjective expected utility $u_{\{\pi\}}$ minus a stochastic cost of ambiguity aversion $\epsilon(u_{\{\pi\}} - u_J)$. If we interpret ϵ as the magnitude of ambiguity aversion, then choice is stochastic due to random ambiguity aversion, while information is fixed. ◊

Example 2. Suppose K is a singleton almost surely or, in other words, $K = \{\pi\}$, where π is a random belief in ΔS. In this case, the maxmin expected utility of p,

$$U_K(p) = \int_H \pi \cdot (u \circ f)\, dp = \pi \cdot (u \circ f_p),$$

is just the subjective expected utility of $f_p := \mathbb{E}_p[f]$, the average act under p. As a result,

$$\rho_A(p) = \mu\big(\{\pi \in \Delta S : \pi \cdot (u \circ f_p) \geq \pi \cdot (u \circ f_q) \text{ for all } q \in A\}\big).$$

In this example, choice is stochastic due to information, while the agent is ambiguity-neutral. Note that this corresponds to the information representation of Lu (2016). Corollary 2 in Section 4 provides a characterization of this special case. ◊

[7] It was even suggested by Ellsberg (1961) as a simple functional form to address his namesake paradox. See Gajdos et al. (2008) and Kopylov (2009) for axiomatic characterizations of deterministic models of ϵ-contamination.

2.2 *Identifying random ambiguity*

We now consider the uniqueness properties of the random ambiguity model. Recall that an act is constant if it yields the same payoff in every state (e.g., a safe asset such as a bond). Theorem 1 below shows that it is possible to identify the utility u and the distribution μ using binary choices where one of the options is a degenerate constant act.

Theorem 1. *Let ρ and τ be represented by (μ, u) and (ν, v) respectively. Then the following statements are equivalent:*

(i) We have $\rho(\delta_f, p) = \tau(\delta_f, p)$ for all constant $f \in H$ and $p \in \Delta H$.

(ii) We have $\mu = \nu$ and $u = \alpha v + \beta$ for $\alpha > 0$.

We now provide a brief outline of the proof. Since identifying u is straightforward from constant acts, we focus on identifying μ. First, note that we can rewrite maxmin expected utility as

$$u_K(f) = 1 - \sigma_K(\mathbf{1} - u \circ f),$$

where $\sigma_K(w) := \max_{\pi \in K} \pi \cdot w$ is the *support function* of K evaluated at $w \in \mathbb{R}^S$. Support functions are thus the dual of maxmin expected utility and identifying μ is equivalent to identifying the distribution of support functions. In the model, comparing a lottery p with a constant act δ_f corresponds to comparing the expectation $p \cdot \sigma_K$ and a constant. Thus, pinning down all such binary choices is equivalent to pinning down the distribution of all expectations of support functions. By using the Stone–Weierstrass theorem, we prove an infinite-dimensional version of the Cramér–Wold theorem to show that the distribution of support functions is uniquely determined by the distribution of their expectations. Finally, uniqueness follows from the one-to-one mapping between a set K and its support function σ_K. The following example provides an illustration.

Example 3. Suppose there are two states $S = \{s_1, s_2\}$ and $X = \{0, 1\}$, and assume $u(x) = x$ for simplicity. Now any set of priors is identified with two beliefs $\underline{\pi}, \bar{\pi} \in [0, 1]$ (about s_1) such that

$$K = \{\pi \in [0, 1] : \underline{\pi} \leq \pi \leq \bar{\pi}\}.$$

Let g be the act that yields $(1, 0)$ (i.e., 1 in state s_1 and 0 in state s_2), let h be the act that yields $(0, 1)$, and let $p = a\delta_g \oplus (1 - a)\delta_h$. For instance, g is an asset that only pays out in the first state, h is an asset that only pays out in the second state, and p is a lottery between the two assets. Note that randomization here is ex ante and occurs before the state is realized. Let f be a constant act that yields c for sure (e.g., yields 1 with probability c and 0 otherwise). For instance, f could be a safe asset. In this case, the probability of choosing the safe asset over the lottery is given by

$$\begin{aligned}\rho(\delta_f, p) &= \mu(\{K \in \mathcal{K} : U_K(\delta_f) \geq U_K(p)\}) \\ &= \mu(\{K \in \mathcal{K} : u_K(f) \geq a u_K(g) + (1 - a) u_K(h)\}) \\ &= \mathbb{P}(\{c \geq a\underline{\pi} + (1 - a)\bar{\pi}\}).\end{aligned}$$

Since $\underline{\pi}$ and $\bar{\pi}$ are positive, their joint distribution is uniquely determined by the distribution of their affine combinations. ◊

The above example illustrates how the richness of ex ante randomization is leveraged for identification. This is similar to identification results involving binary menus in other domains, such as Lu (2016) and Lu and Saito (2018, 2020). The last paper also proves an infinite-dimensional version of the Cramér–Wold theorem; while they achieve this by focusing on the compact set of Lipschitz continuous functions, we focus on the compact set of support functions.

When the richness of ex ante randomization cannot be leveraged, identification requires richer choice data. For instance, in the exercise in Example 3, restricting choice options to degenerate acts identifies only the marginal distributions of $\underline{\pi}$ and $\bar{\pi}$, but not their joint distribution. Intuitively, ex ante lotteries allow us to evaluate the entire support function, while in the absence of ex ante lotteries, support functions can only be evaluated at two points in binary comparisons. Identification, however, can be achieved by looking at stochastic choice over all menus (see Theorem 7 in the Appendix); this is possible despite the well known difficulties in identifying models of random nonexpected utility (see Lin 2020).

2.3 *Comparing stochastic ambiguity*

We now study comparative statics of the random ambiguity model. We introduce a new stochastic order over sets that corresponds to greater uncertainty aversion under stochastic choice. One stochastic choice is *more uncertainty-averse than* another if constant acts are chosen more frequently in the former than the latter.

Definition 3. Stochastic choice ρ is *more uncertainty-averse than* τ if $\rho_A(\delta_f) \geq \tau_A(\delta_f)$ for all constant $f \in H$ and $A \in \mathcal{A}$.

We now introduce a new stochastic order over sets. A family of sets of beliefs $\mathcal{J} \subset \mathcal{K}$ is *increasing* if $K \supset J \in \mathcal{J}$ implies $K \in \mathcal{J}$. A distribution μ of sets *stochastically c-dominates (or convex-dominates)* another distribution ν if μ puts greater weight than ν on all increasing, convex, and closed families of sets.

Definition 4. Distribution μ *stochastically c-dominates* ν, that is $\mu \geq_c \nu$, if $\mu(\mathcal{J}) \geq \nu(\mathcal{J})$ for all increasing, convex, and closed $\mathcal{J} \subset \mathcal{K}$.

The standard stochastic order comparing sets asserts that μ stochastically dominates ν if μ puts greater weight than ν on all increasing and closed families of sets, i.e., $\mu(\mathcal{J}) \geq \nu(\mathcal{J})$ for all increasing $\mathcal{J} \subset \mathcal{K}$. This is first-order stochastic dominance with respect to set inclusion and is equivalent to the existence of two random sets K_1 and K_2 with marginal distributions or laws μ and ν, respectively, such that $K_2 \subset K_1$ almost surely. This equivalence is known as Strassen's theorem (see Theorem 4.41 of Molchanov 2005). The c-dominance stochastic order further requires that the families of sets have to be convex. Our notion of stochastic c-dominance is thus weaker than the standard stochastic dominance as the following example illustrates.

EXAMPLE 4. Suppose there are two states $S = \{s_1, s_2\}$ and let π_i be the degenerate belief on state s_i for $i \in \{1, 2\}$. Let $\mu = \frac{1}{2}\delta_{\{\frac{1}{2}\pi_1+\frac{1}{2}\pi_2\}} + \frac{1}{2}\delta_{\Delta S}$ and $\nu = \frac{1}{2}\delta_{\{\pi_1\}} + \frac{1}{2}\delta_{\{\pi_2\}}$. It is straightforward to show that $\mu \geq_c \nu$. To see why μ does not dominate ν in the standard stochastic ordering, consider

$$\mathcal{J} := \{K \in \mathcal{K} : \pi_1 \in K \text{ or } \pi_2 \in K\}.$$

Note that $\mathcal{J}$ is closed and increasing, but not convex (e.g., $\{\pi_1\}, \{\pi_2\} \in \mathcal{J}$, but $\{\frac{1}{2}\pi_1 + \frac{1}{2}\pi_2\} \notin \mathcal{J}$). Since $\mu(\mathcal{J}) = \frac{1}{2} < 1 = \nu(\mathcal{J})$, μ does not dominate ν in the standard stochastic order. ◊

The following result shows that greater uncertainty aversion is exactly characterized by c-dominance.

THEOREM 2. *Let ρ and τ be represented by (μ, u) and (ν, v) respectively. Then ρ is more uncertainty-averse than τ if and only if $\mu \geq_c \nu$ and $u = \alpha v + \beta$ for $\alpha > 0$.*

Under deterministic maxmin expected utility, more uncertainty aversion corresponds to having a larger set of priors. Extending to stochastic choice, one may expect this to correspond to having a larger set of priors almost surely, i.e., the standard stochastic dominance for sets. Theorem 2 shows that this is not the case. Consider Example 4 and suppose ρ has distribution $\mu = \frac{1}{2}\delta_{\{\frac{1}{2}\pi_1+\frac{1}{2}\pi_2\}} + \frac{1}{2}\delta_{\Delta S}$, while τ has distribution $\nu = \frac{1}{2}\delta_{\{\pi_1\}} + \frac{1}{2}\delta_{\{\pi_2\}}$. Suppose both have the same utility u and consider some act f. If $u(f(s_2)) \geq u(f(s_1))$, then

$$u_{\{\pi_1\}}(f) = u\big(f(s_1)\big) \geq \min\big\{u\big(f(s_1)\big), u\big(f(s_2)\big)\big\} = u_{\Delta S}(f)$$

$$u_{\{\pi_2\}}(f) = u\big(f(s_2)\big) \geq \frac{1}{2}u\big(f(s_1)\big) + \frac{1}{2}u\big(f(s_2)\big) = u_{\{\frac{1}{2}\pi_1+\frac{1}{2}\pi_2\}}(f).$$

The case for $u(f(s_2)) \leq u(f(s_1))$ is symmetric, so maxmin expected utility under ν first-order stochastically dominates maxmin expected utility under μ. This argument easily extends to menus of lotteries, so ρ is more uncertainty-averse than τ despite μ not dominating ν in the standard stochastic order (see Example 4).

To understand the gap between greater uncertainty aversion and the standard stochastic order over sets, suppose ρ and τ reflect the stochastic choice of two populations. If, for every agent in the first group, there is a corresponding agent in the second group with a smaller set of priors, then the first group is obviously more uncertainty-averse than the second group. The converse, however, is not true as demonstrated above. While the agent with the set of priors ΔS is larger than both $\{\pi_1\}$ and $\{\pi_2\}$, the agent with the singleton set $\{\frac{1}{2}\pi_1 + \frac{1}{2}\pi_2\}$ is larger than neither. Nevertheless, his evaluation of an act is always lower than that of either $\{\pi_1\}$ or $\{\pi_2\}$, so ρ is more uncertainty-averse than τ.

When $\mu = \delta_K$ and $\nu = \delta_J$ are degenerate, c-dominance and standard dominance coincide, and both are equivalent to $K \subset J$. Theorem 2 is a generalization of corresponding results under deterministic choice (see Theorem 17(ii) of Ghirardato and Marinacci

2002). Note that stochastic choice allows for a bit more flexibility and subtlety in comparing random sets. As in the deterministic counterpart, the order is incomplete, as some random sets are incomparable.

Our characterization of greater uncertainty aversion relies on the richness of ex ante randomization. In the absence of ex ante lotteries, characterizing greater uncertainty aversion leads to a different (weaker) stochastic order than c-dominance. This would allow for more comparisons between random sets and lead to coarser comparative statics.[8]

2.4 *Updating and dynamic consistency*

The random ambiguity model is silent on the interpretation of priors and updating rules. In the classic model of information where the random set is always a singleton belief (Example 2), μ is simply a distribution over posterior beliefs. Under Bayesian updating, the prior is the expectation of posterior beliefs. We now extend this definition to random sets of posterior beliefs and consider updating.

Let ν be a distribution on ΔS. A random element taking values in ΔS with distribution ν is called a *random belief*. Note that a random belief is a special case of a random set where the set takes on only singleton values, i.e., $K = \{\pi\}$ almost surely. Random beliefs correspond to the classic way to model information. If ν is the distribution of a random belief, then Bayes' rule dictates that its expectation

$$\mathbb{E}_\nu[\pi] := \int_{\Delta S} \pi \, d\nu \tag{1}$$

is exactly the prior $\mathbb{E}_\nu[\pi] \in \Delta S$.

We call a random belief a *posterior selection* of a random set if it is contained in the random set almost surely. The formal definition is as follows.

Definition 5. Let π be a random belief with distribution ν and let K be a random set with distribution μ. Then π is a *posterior selection* of K if there exists a probability law η on $\Delta S \times \mathcal{K}$ such that the following statements hold:

(i) The marginals of η on ΔS and $\mathcal{K}$ are ν and μ, respectively.

(ii) We have $\pi \in K$ almost surely with respect to η.

Given a random set with distribution μ, let $\Pi(\mu)$ denote the family of all distributions of its posterior selections. In other words, if a random belief with distribution ν is a posterior selection of a random set with distribution μ, then $\nu \in \Pi(\mu)$.

Consider a random set with distribution μ. Since every posterior selection is a random belief, we can calculate its prior $\mathbb{E}_\nu[\pi]$. The *prior set* of μ is the family of priors of all posterior selections of the random set. This is known as the Aumann (1965) expectation of the random set.

[8]Consider Example 4, but now let $\nu = \frac{1}{2}\delta_{K_1} + \frac{1}{2}\delta_{K_2}$, where K_i for $i \in \{1, 2\}$ is the set of priors where the belief on s_i is greater than $\frac{1}{2}$. In this case, μ dominates ν in this weaker order, but $\mu \not\geq_c \nu$.

DEFINITION 6. The *prior set* of μ is denoted by

$$\mathbb{E}_\mu[K] := \{\mathbb{E}_\nu[\pi] : \nu \in \Pi(\mu)\}.$$

It is straightforward to see that $\mathbb{E}_\mu[K]$ is convex and compact, so $\mathbb{E}_\mu[K] \in \mathcal{K}$.[9] We can also write the prior set as

$$\mathbb{E}_\mu[K] = \int_{\mathcal{K}} K\, d\mu, \tag{2}$$

where the integral is taken as the limit of Minkowski averages. This is known as the Debreu (1967) expectation and is equivalent to the Aumann expectation for random compact convex sets in Euclidean space (see Theorem 1.26 of Molchanov 2005). Note that (2) is the generalization of Bayes' rule from (1) to sets. When the agent is ambiguity-neutral so that the random set is a random belief, the prior set $\mathbb{E}_\mu[K] = \{\mathbb{E}_\nu[\pi]\}$ coincides with the singleton prior as in the standard Bayesian model of information.

What is a natural interpretation of the prior set being the Aumann expectation of a random set of posteriors? The following example provides an illustration.

EXAMPLE 5. Suppose there are two states $S = \{s_1, s_2\}$ and two possible signal realizations $\Theta = \{\theta_1, \theta_2\}$. A signal structure is a joint distribution on $S \times \Theta$. We can express this as a matrix $T \in \mathbb{R}^{S\times\Theta}$, where T_{ij} is the probability of (s_i, θ_j) for $i, j \in \{1, 2\}$. Consider the four signal structures

$$T_1 = \begin{pmatrix} \frac{1}{2} & \frac{1}{8} \\ 0 & \frac{3}{8} \end{pmatrix} \qquad T_2 = \begin{pmatrix} \frac{1}{2} & \frac{3}{8} \\ 0 & \frac{1}{8} \end{pmatrix} \qquad T_3 = \begin{pmatrix} 0 & \frac{1}{8} \\ \frac{1}{2} & \frac{3}{8} \end{pmatrix} \qquad T_4 = \begin{pmatrix} 0 & \frac{3}{8} \\ \frac{1}{2} & \frac{1}{8} \end{pmatrix}.$$

Suppose the agent believes the true signal structure can be any convex combination of these four. Note that while the prior on S may vary across the signal structures, the distribution of signal realizations on Θ is always 50–50. Now, conditional on θ_1, the agent's belief (about s_1) is in the entire set $[0, 1]$, while conditional on θ_2, his belief is in the set $[\frac{1}{4}, \frac{3}{4}]$. Thus, $\mu = \frac{1}{2}\delta_{[0,1]} + \frac{1}{2}\delta_{[\frac{1}{4},\frac{3}{4}]}$. The set of possible priors is $[\frac{1}{8}, \frac{7}{8}]$. From (2), this is exactly equal to

$$\mathbb{E}_\mu[K] = \frac{1}{2}[0, 1] + \frac{1}{2}\left[\frac{1}{4}, \frac{3}{4}\right],$$

which is the Aumann expectation. $\Diamond$

Example 5 provides a natural setup where a set of priors is exactly the Aumann expectation of a random set. We leave the interpretation of signal structures open; they can be purely subjective to the agent or they can correspond to objective pieces of information. Note that they all generate the same distribution of signal realizations (in Θ); this is consistent with a repeated choice interpretation of stochastic choice where the agent eventually learns the true signal distribution.

[9]This follows from the fact that $\Pi(\mu)$ is a convex compact set (Proposition 2.21 of Molchanov 2005).

In the example, if we considered only signal structures that are combinations of T_2 and T_3, then the analysis would be the same. This is not true if we considered only combinations of T_1 and T_4; μ would still be $\frac{1}{2}\delta_{[0,1]} + \frac{1}{2}\delta_{[\frac{1}{4},\frac{3}{4}]}$, but the set of possible priors would be $[\frac{3}{8}, \frac{5}{8}]$, which is a strict subset of the Aumann expectation. This is because while combinations of T_2 and T_3 and combinations of T_1 and T_4 both generate the same μ, the former is more uncertain from an ex ante perspective, as it is consistent with a larger set of possible priors. The Aumann expectation corresponds to the largest set of possible priors for all signal structures consistent with μ.

We now show that interpreting the prior set as the Aumann expectation of a random set of beliefs satisfies dynamic consistency. Consider a preference relation $\succeq$ on ΔH. We interpret $\succeq$ as representing ex ante preferences before the agent receives any information. The agent satisfies *dynamic consistency* if his ex ante preference is consistent with his stochastic choice.

Definition 7. Preference relation $\succeq$ and stochastic choice ρ satisfy *dynamic consistency* if $\rho(p, q) = 1$ implies $p \succeq q$.

Dynamic consistency implies that whenever the agent chooses p over q for sure, then he prefers p over q ex ante. This is a relatively weak notion of consistency in that it considers only events that occur with probability 1 under stochastic choice. We say that $\succeq$ is represented by (K, u) if it is represented by a maxmin expected utility with a (deterministic) prior set $K \in \mathcal{K}$ and a nonconstant vNM utility u.

Theorem 3. *If $\succeq$ and ρ are represented by $(\mathbb{E}_\mu[K], u)$ and (μ, u), respectively, then they satisfy dynamic consistency.*

Proof. The support function of the Aumann expectation is the expectation of the support function of the random set (see Theorem 1.26 of Molchanov 2005). This implies that

$$U_{\mathbb{E}_\mu[K]}(p) = \int_{\mathcal{K}} U_K(p)\, d\mu.$$

Since $\rho(p, q) = 1$, $U_K(p) \geq U_K(q)$ almost surely, so $U_{\mathbb{E}_\mu[K]}(p) \geq U_{\mathbb{E}_\mu[K]}(q)$. Thus, $p \succeq q$ as desired. $\square$

In this model, ex ante preferences are exactly captured by maxmin expected utility where the set of prior beliefs is given by the Aumann expectation of the random set of posteriors. Since the Aumann expectation preserves linearity with respect to Minkowski mixing, this stochastic model of updating under ambiguity preserves dynamic consistency and consequentialism. Note that updating in this form is not necessarily prior-by-prior and signals are not necessarily partitional.[10] Modeling random sets of beliefs can thus suggest new restrictions on updating behavior.

[10]When signals are partitional, the Aumann expectation satisfies the rectangularity condition of Epstein and Schneider (2003).

3. Characterization

We now provide an axiomatic characterization of the random ambiguity model. First, we leverage ex ante randomization to obtain a random (ex ante) expected utility representation. We then impose conditions to ensure that the random (ex post) Bernoulli utilities are maxmin expected utilities.

The first axiom (Axiom 1) consists of four conditions from Gul and Pesendorfer (2006) with respect to ex ante randomization. *Monotonicity* ensures that the probability an option is chosen is decreasing in the size of the menu and is necessary for any random utility model. *Ex ante independence* is the standard independence axiom applied to ex ante randomization. A lottery $p \in A$ is *(ex ante) extreme* in A if it is not in the interior of the convex hull (with respect to ex ante randomization) of A. Let $\text{ext}^{\oplus}(A)$ denote the set of extreme lotteries of A. *Ex ante extremeness* asserts that ex ante randomization is not helpful, so only extreme lotteries are chosen, barring ties. *Continuity* says that the stochastic choice mapping restricted to the domain $\mathcal{A}^{\circ} \subset \mathcal{A}$ of menus without ties is continuous.

Axiom 1.1 (Monotonicity). *For all $p \in A \subset B$ and $A, B \in \mathcal{A}$, $\rho_A(p) \geq \rho_B(p)$.*

Axiom 1.2 (Ex ante independence). *For all $p \in A \in \mathcal{A}$, $q \in \Delta H$, and $a > 0$, $\rho_A(p) = \rho_{aA \oplus (1-a)q}(ap \oplus (1-a)q)$.*

Axiom 1.3 (Ex ante extremeness). *For all $A \in \mathcal{A}$, $\rho_A(\text{ext}^{\oplus}(A)) = 1$.*

Axiom 1.4 (Continuity). *The mapping $\rho : \mathcal{A}^{\circ} \to \Delta(\Delta H)$ is continuous.*

The next axiom is the stochastic analog to uncertainty aversion and preference for hedging. To illustrate, consider two acts f and g, and the ex post mixture $h = af + (1-a)g$ for $a \in (0, 1)$. Note that h is the hedge option. For deterministic preferences, uncertainty aversion means that h is ranked higher than the worst of f and g. This implies four possible preference rankings:

$$\begin{array}{ll} f \succeq h \succeq g & \quad g \succeq h \succeq f \\ h \succeq f \succeq g & \quad h \succeq g \succeq f. \end{array}$$

Among the four rankings, f is preferred to both g and h if and only if f is preferred to just h. For stochastic preferences, this means that the probability of choosing f over both g and h is the same as the probability of just choosing f over h. In other words, one can ignore the other option g when the hedge option h is available.

Axiom 2 (ex post hedging) generalizes this concept. Removing an option does not affect the choice of another option as long as the hedge option (i.e., mixture of the two options) is available. In Section 4, we show that this axiom characterizes random quasiconcave utility and is exactly the stochastic analog of convex preferences in the deterministic setting. We use the notation δ_F to denote the menu of degenerate acts

$$\delta_F := \{\delta_f : f \in F\}$$

for any finite set of acts $F \subset H$.

Axiom 2 (Ex post hedging). *For any $f \in F$ and $g \in H$, if $af + (1-a)g \in F$, then $\rho_{\delta_F}(\delta_f) = \rho_{\delta_{F \cup \{g\}}}(\delta_f)$.*

An interesting implication of ex post hedging combined with monotonicity is that an option is chosen less frequently when other options are "closer" to it via randomization. To illustrate, consider $h = af + (1-a)g$ as before, so the act f is closer to h than it is to g. If we let $F = \{f, h, g\}$, then by ex post hedging and monotonicity,

$$\rho(\delta_f, \delta_h) = \rho_{\delta_F}(\delta_f) \le \rho(\delta_f, \delta_g).$$

Thus, the act f is chosen less frequently when other acts are closer. This is the stochastic analog of uncertainty aversion under deterministic choice, i.e., $f \succ h$ implies $f \succ g$. Note, however, that this is weaker than ex post hedging and is insufficient to ensure the random utility is quasiconcave.

The next axiom relaxes reversal of order and states that the agent does not distinguish between ex ante and ex post randomization when mixing with constant acts. To illustrate, recall that under maxmin expected utility, deterministic preferences satisfy the independence axiom when mixed with constant acts (the certainty-independence axiom of Gilboa and Schmeidler 1989). In other words, if δ_f is indifferent to δ_g, then $\delta_{af+(1-a)h}$ is also indifferent to $\delta_{ag+(1-a)h}$ for any constant h. Suppose g is also constant and the agent is indifferent to timing whenever there is no uncertainty. Since constant acts convey no uncertainty, $\delta_{ag+(1-a)h}$ is indifferent to $a\delta_g \oplus (1-a)\delta_h$. Finally, since the agent satisfies the independence axiom for ex ante randomization,

$$a\delta_f \oplus (1-a)\delta_h \sim a\delta_g \oplus (1-a)\delta_h \sim \delta_{af+(1-a)h}.$$

Axiom 3 (certainty reversal of order) extends this reasoning for stochastic choice. It states that ex post and ex ante randomization with a constant act are interchangeable in binary comparisons.

Axiom 3 (Certainty reversal of order). *If $h \in H$ is constant, then for any $f \in H$ and $p \in \Delta H$,*

$$\rho\big(a\delta_f \oplus (1-a)\delta_h, p\big) = \rho(\delta_{af+(1-a)h}, p).$$

The last three axioms are conditions from the information representation of Lu (2016). Axiom 4 (certainty determinism) states that choice is deterministic over menus consisting of constant acts. It ensures that all stochasticity is due to ambiguity attitudes or information. We say the menu δ_F is *constant* if F consists only of constant acts.

Axiom 4 (Certainty determinism). *For all constant δ_F, $\rho_{\delta_F}(\delta_f) \in \{0, 1\}$.*

Axiom 5 (dominance) states that if an act is the best regardless of which state occurs, then it must be chosen for sure. It is the stochastic analog of the standard state monotonicity axiom necessary for maxmin expected utility. Given $f \in H$ and $s \in S$, let $f_s \in H_c$

denote the constant act such that $f_s(s') = f(s)$ for all $s' \in S$. Given finite $F \subset H$, let

$$F_s := \{f_s : f \in F\}$$

denote the set of constant acts in state $s \in S$. Note that the menu δ_{F_s} is constant.

Axiom 5 (Dominance). *If $\rho_{\delta_{F_s}}(\delta_{f_s}) = 1$ for all $s \in S$, then $\rho_{\delta_F}(\delta_f) = 1$ for any $f \in F$.*

Finally, nondegeneracy rules out the trivial case of universal indifference.

Axiom 6 (Nondegeneracy). *There exists F and some $f \in F$ such that $\rho_{\delta_F}(\delta_f) < 1$.*

We are now ready to present the main representation result.

Theorem 4. *Stochastic choice ρ satisfies Axioms 1–6 if and only if it is represented by (μ, u).*

The primary difficulty of the proof lies in the construction of a random utility representation. Incorporating ex ante randomization in our setup allows us to leverage the random expected utility characterization of Gul and Pesendorfer (2006). Intuitively, linearity lowers the dimensionality of the space of possible utilities, allowing for a simpler axiomatization. In the absence of ex ante randomization, a clean and intuitive characterization of random nonlinear utility would be difficult without imposing additional parametric restrictions.[11]

The Gul and Pesendorfer (2006) characterization involves lotteries over a finite-dimensional set, while we have lotteries over the set of all acts, an infinite-dimensional set. Lu and Saito (2020) provide an infinite-dimensional extension of random expected utility, but the extension applies when utilities are Lipschitz continuous with common bound. The main insight of the proof is that our axioms are sufficient for Lipschitz continuity (see Step 3 in Appendix A.4). Once a random utility representation is established, it is straightforward to show that our axioms imply that the random utility must be maxmin expected utility almost surely. It is worth pointing out that this approach can be used to obtain characterizations of other ambiguity representations beyond maxmin expected utility (e.g., a random Choquet expected utility).

4. Stochastic ambiguity attitudes

In this section, we study the general stochastic properties of nonlinearity in models of random utility. For simplicity, we focus exclusively on ex post randomization in this section.

In the random ambiguity model, Bernoulli utilities are maxmin expected utilities and exhibit ambiguity aversion. As a result, ex post randomization is desirable and utilities are quasiconcave. We show that for random utilities, quasiconcavity is exactly characterized by ex post hedging. Alternatively, if the agent is ambiguity-loving, then ex post

[11] Of course, one can always directly impose the Block–Marschak inequalities.

randomization is not desirable and utilities are quasiconvex. We show that for random utilities, quasiconvexity is exactly characterized by the ex post version of the extremeness axiom (Axiom 1.3). Our results thus provide a theoretical foundation for studying other stochastic ambiguity models beyond random maxmin expected utility.

We say a stochastic choice is ex post hedging if it satisfies Axiom 2 above.

Definition 8. Stochastic choice ρ is *ex post hedging* if $af + (1 - a)g \in F$ implies $\rho_{\delta_F}(\delta_f) = \rho_{\delta_{F\cup\{g\}}}(\delta_f)$ for any $f \in F$ and $g \in H$.

We can weaken ex post hedging to ex post mixture-loving, which applies only to menus that contain two acts and their mixture. Note that when there is a one-dimensional ordering, this is exactly the strong centrality axiom of Apesteguia et al. (2017) that characterizes single-peakedness.

Definition 9. Stochastic choice ρ is *ex post mixture-loving* if $h = af + (1 - a)g$ implies $\rho_{\delta_{\{f,h\}}}(\delta_f) = \rho_{\delta_{\{f,h,g\}}}(\delta_f)$ for any $f, g \in H$.

Alternatively, ex post randomization would not be attractive for an ambiguity-loving agent. As a result, ex post mixtures of acts are never chosen in any menu. An act $f \in F$ is (ex post) extreme in a finite $F \subset H$ if it is not in the interior of the convex hull of F. Let $\text{ext}(F)$ denote the set of extreme acts of F. Ex post extremeness asserts that only extreme acts are chosen, barring ties.

Definition 10. Stochastic choice ρ is *ex post extreme* if $\rho_{\delta_F}(\delta_{\text{ext}(F)}) = 1$ for all finite $F \subset H$.

We can weaken ex post extremeness to the following model, which applies only to menus that contain two acts and their mixture.

Definition 11. Stochastic choice ρ is *ex post mixture-averse* if $\rho_{\delta_{\{f,af+(1-a)g,g\}}}(\{\delta_f, \delta_g\}) = 1$ for all $f, g \in H$.

We now consider a general random utility model that may be ambiguity-averse or ambiguity-loving. Let V be the set of continuous utilities $v : H \to \mathbb{R}$ and let ν be a probability measure on V. We say that ρ is *represented by* ν if

$$\rho_{\delta_F}(\delta_f) = \nu\big(\{v \in V : v(f) \geq v(g) \text{ for all } g \in F\}\big).$$

For instance, in the random ambiguity model, ν is a distribution on maxmin expected utilities. We can now define quasiconcave and quasiconvex as stochastic properties of the random utility.

Definition 12. The utility distribution ν has the following properties:

(i) It is *quasiconcave* if $v(f) \geq v(g)$ implies $v(af + (1 - a)g) \geq v(g)$ almost surely.

(ii) It is *quasiconvex* if $v(f) \geq v(g)$ implies $v(f) \geq v(af + (1 - a)g)$ almost surely.

The main result of this section shows that random quasiconcave utilities are characterized by ex post hedging (or mixture loving), while random quasiconvex utilities are characterized by ex post extremeness (or mixture aversion). For simplicity, assume ν excludes ties, i.e., $v(f) = v(g)$ occurs with zero probability.[12]

Theorem 5. *Let ρ be represented by ν that excludes ties.*

(i) Distribution ν is quasiconcave if and only if ρ is ex post mixture-loving if and only if ρ is ex post hedging.

(ii) Distribution ν is quasiconvex if and only if ρ is ex post mixture-averse if and only if ρ is ex post extreme.

Theorem 5 states that ex post hedging (or mixture loving) and ex post extremeness (or mixture aversion) are exactly the stochastic properties that characterize quasiconcavity and quasiconvexity, respectively, for random utilities. Hedging and extremeness are dual properties: while the former focuses on hedge options, the latter focuses on extreme options. Recall that for deterministic preferences, a utility is quasiconcave if and only if the preference relation it represents is convex.[13] Thus, ex post hedging (or mixture loving) is exactly the stochastic choice analog of convex preferences.

When stochastic choice is both ex post mixture-loving and -averse, we say it is *ex post mixture-neutral*. A random utility ν satisfies betweenness if it is both quasiconcave and quasiconvex, i.e., the random utilities satisfy betweenness almost surely.[14] The following characterization of random betweenness is immediate from Theorem 5.

Corollary 1. *Let ρ be represented by ν that exclude ties. Then ν satisfies betweenness if and only if ρ is ex post mixture-neutral.*

The most well known case of a random utility that satisfies betweenness is random expected utility. While an ex post version of the independence axiom (Axiom 1.2) is necessary for random expected utility, it is not sufficient. In fact, Lin (2020) provides an example of a random nonexpected utility that satisfies ex post independence. Under random utility and ex post independence, it is straightforward to show that ex post hedging and ex post extremeness are equivalent. This suggests an alternate axiomatization of random expected utility where we substitute the hedging axiom for the extremeness axiom of Gul and Pesendorfer (2006).

Finally, we return to our random ambiguity model. It is easy to see that ex post extremeness is sufficient to ensure that all random utilities are subjective expected utilities. This provides a simple characterization of the special case where the agent is ambiguity-neutral almost surely and choice is stochastic only due to information (see Example 2).

---

[12]If we allow for ties to occur with probability 1 as in regular utility distributions, then ex post hedging is equivalent to a slightly stronger statement of quasiconcavity. See Theorem 6 in the Appendix for details.

[13]That is, $f \succeq g$ implies $af + (1-a)g \succeq g$ for all $a \in [0, 1]$.

[14]Betweenness means that $\min\{v(f), v(g)\} \leq v(af + (1-a)g) \leq \max\{v(f), v(g)\}$ for all $a \in [0, 1]$. Dekel (1986) provides an axiomatic characterization of betweenness preferences.

Corollary 2. *Suppose ρ is represented by (μ, u) and excludes ties. Then ρ is ex post extreme if and only if K is a singleton almost surely.*

The proof follows from Theorem 5 and the fact that a maxmin expected utility that satisfies betweenness is linear.

The results in this section provide a foundation for the study of more general models of stochastic ambiguity. Although we focused only on ex post randomization in this section, our results on the stochastic properties of random nonlinear utilities extend to any mixture space more generally.

Appendix

A.1 *Proof of Theorem 1*

Let ρ and τ be represented by (μ, u) and (ν, v), respectively. That (ii) implies (i) is straightforward. We show that (i) implies (ii). Suppose (i) is true and note that for constant $f, g \in H_c$, $u(f) \geq u(g)$ if and only if $\rho(\delta_f, \delta_g) = 1$ if and only if $\tau(\delta_f, \delta_g) = 1$ if and only if $v(f) \geq v(g)$. Thus, $u = \alpha v + \beta$ for $\alpha > 0$. Without loss of generality, assume $u = v$ and $1 = u(\bar{x}) \geq u(x) \geq u(\underline{x}) = 0$ for $\bar{x}, \underline{x} \in X$ and all $x \in X$.

Let $W := [0, 1]^n$, where $n = |S|$. Define the mapping $\psi : H \to W$ such that $\psi(h) = \mathbf{1} - u \circ h$. Note that

$$
\begin{aligned}
U_K(p) &= \int_H u_K(h)\, dp = \int_H \min_{\pi \in K} \pi \cdot (u \circ h)\, dp \\
&= \int_W \Big(1 - \max_{\pi \in K} \pi \cdot w\Big)\, dr = 1 - \sigma_K \cdot r,
\end{aligned}
$$

where $r = p \circ \varphi^{-1} \in \Delta W$ and σ_K is the support function of $K \in \mathcal{K}$. Let $f = (1-\alpha)\bar{x} + \alpha \underline{x}$, which is constant. Thus,

$$
\begin{aligned}
\rho(\delta_f, p) &= \mu\big(\big\{K \in \mathcal{K} : u(f) \geq U_K(p)\big\}\big) \\
&= \mu\big(\{K \in \mathcal{K} : 1 - \alpha \geq 1 - \sigma_K \cdot r\}\big) \\
&= \mu\big(\{K \in \mathcal{K} : \sigma_K \cdot r \geq \alpha\}\big).
\end{aligned}
$$

Since $\rho(\delta_f, p) = \tau(\delta_f, p)$, $\sigma_K \cdot r$ has the same distribution under μ and ν for all $r \in \Delta W$.

Let Σ denote the set of support functions σ_K and let $C(\Sigma)$ denote the set of continuous functions on Σ. Let $\Phi \subset C(\Sigma)$ denote the set of functions

$$
\phi(\sigma_K) = \sum_i a_i e^{\lambda_i \sigma_K \cdot r_i}
$$

for $a_i \in \mathbb{R}$, $\lambda_i \geq 0$, and $r_i \in \Delta W$. Since each $\sigma_K \cdot r_i$ has the same distribution under μ and ν,

$$
\int_{\mathcal{K}} \phi(\sigma_K)\, d\mu = \int_{\mathcal{K}} \sum_i a_i e^{\lambda_i \sigma_K \cdot r_i}\, d\mu = \int_{\mathcal{K}} \sum_i a_i e^{\lambda_i \sigma_K \cdot r_i}\, d\nu = \int_{\mathcal{K}} \phi(u)\, d\nu.
$$

We show that Φ is dense in $C(\Sigma)$. First note that Φ is a vector space that includes constants, since $e^{0\sigma_K \cdot r} = 1 \in \Phi$. Consider $a_1 e^{\lambda_1 \sigma_K \cdot r_1}, a_2 e^{\lambda_2 \sigma_K \cdot r_2} \in \Phi$. If $\lambda_1 + \lambda_2 > 0$, then

$$a_1 e^{\lambda_1 \sigma_K \cdot r_1} a_2 e^{\lambda_2 \sigma_K \cdot r_2} = a_1 a_2 e^{(\lambda_1+\lambda_2)\sigma_K \cdot (\frac{\lambda_1}{\lambda_1+\lambda_2} r_1 + (1-\frac{\lambda_1}{\lambda_1+\lambda_2}) r_2)} \in \Phi.$$

Alternatively, if $\lambda_1 + \lambda_2 = 0$, then $\lambda_1 = \lambda_2 = 0$ and

$$a_1 e^{\lambda_1 \sigma_K \cdot r_1} a_2 e^{\lambda_2 \sigma_K \cdot r_2} = a_1 a_2 \in \Phi.$$

This means that Φ is closed under multiplication.

Next we show that Φ separates points in Σ. Suppose $\sigma_K, \sigma_J \in \Sigma$ such that $\sigma_K \neq \sigma_J$. Thus, there is some $w \in W$ such that $\sigma_K(w) > \sigma_J(w)$ without loss of generality. If we let $r = \delta_w$, then $\sigma_K \cdot r = \sigma_K(w) > \sigma_J(w) = \sigma_J \cdot r$, so $e^{\sigma_K \cdot r} > e^{\sigma_J \cdot r}$. This establishes that Φ separates points in Σ.

Since $\mathcal{K}$ is compact (Theorem 3.85 of Aliprantis and Border 2006; henceforth AB), the homeomorphism between Σ and $\mathcal{K}$ implies that Σ is also compact. Since Φ is a subalgebra, contains the constant function, and separates points in Σ, Φ is uniformly dense in $C(\Sigma)$ by the Stone–Weierstrass theorem (Theorem 9.13 of AB). This means that for any $\phi \in C(\Sigma)$, we can find $\phi_i \in \Phi$ such that $\phi_i \to \phi$ uniformly. By dominated convergence,

$$\int_{\mathcal{K}} \phi(\sigma_K)\, d\mu = \lim_i \int_{\mathcal{K}} \phi_i(\sigma_K)\, d\mu = \lim_i \int_{\mathcal{K}} \phi_i(\sigma_K)\, d\nu = \int_{\mathcal{K}} \phi(\sigma_K)\, d\nu.$$

By Theorem 15.1 of AB, support functions have the same distribution under μ and ν. In other words, for any $J \in \mathcal{K}$,

$$\mu\big(\{K \in \mathcal{K} : \sigma_K \leq \sigma_J\}\big) = \nu\big(\{K \in \mathcal{K} : \sigma_K \leq \sigma_J\}\big)$$
$$\mu\big(\{K \in \mathcal{K} : K \subset J\}\big) = \nu\big(\{K \in \mathcal{K} : K \subset J\}\big).$$

Since containment functionals completely characterize the distribution of convex compact sets (Theorem 7.8 of Molchanov 2005), $\mu = \nu$ as desired.

A.2 *Proof of Theorem 2*

Suppose ρ is more uncertainty-averse than τ. Note that if $u \neq \alpha v + \beta$ for $\alpha > 0$, then we can find two constant acts $f, g \in H$ such that $\tau(\delta_f, \delta_g) = 1 > 0 = \rho(\delta_f, \delta_g)$, which contradicts the fact that ρ is more uncertainty-averse than τ. Hence, we can assume $u = v$ without loss of generality, and $1 = u(\bar{x}) \geq u(x) \geq u(\underline{x}) = 0$ for $\bar{x}, \underline{x} \in X$ and all $x \in X$.

Let $C(W)$ denote the set of all continuous functions on the compact set $W := [0,1]^n$, where $n = |S|$. Note that we can define the dual pair

$$\varphi \cdot r := \int_W \varphi(w)\, dr,$$

where $\varphi \in C(W)$ and r is a signed Borel measure on W (see Corollary 14.15 of AB). Fix some increasing, closed, and convex $\mathcal{J}$, and define

$$C_{\mathcal{J}} := \big\{\varphi \in C(W) : \varphi \geq \sigma_J \text{ for some } J \in \mathcal{J}\big\}.$$

First, we prove two claims about $C_{\mathcal{J}}$.

Claim 1. *A set $K \in \mathcal{J}$ if and only if $\sigma_K \in C_{\mathcal{J}}$.*

Proof. If $K \in \mathcal{J}$, then $\sigma_K \in C_{\mathcal{J}}$ trivially. Suppose $\sigma_K \in C_{\mathcal{J}}$ so $\sigma_K \geq \sigma_J$ for some $J \in \mathcal{J}$. Thus, $K \supset J$ and since $\mathcal{J}$ is increasing, $K \in \mathcal{J}$. □

Claim 2. *The set $C_{\mathcal{J}}$ is closed and convex.*

Proof. We first show that $C_{\mathcal{J}}$ is closed. Consider $\varphi_k \in C_{\mathcal{J}}$ such that $\varphi_k \to \varphi$. Thus, $\varphi_k \geq \sigma_{J_k}$ for $J_k \in \mathcal{J}$. Since $\mathcal{J} \subset \mathcal{K}$ is closed and $\mathcal{K}$ is compact, $\mathcal{J}$ is also compact. Thus, we can find a convergent subsequence $J_i \to J \in \mathcal{J}$ and $\sigma_{J_i} \to \sigma_J$. Thus, $\varphi \geq \sigma_J$, so $\varphi \in C_{\mathcal{J}}$.

We now show that $C_{\mathcal{J}}$ is convex. Suppose $\varphi_1, \varphi_2 \in C_{\mathcal{J}}$, so $\varphi_i \geq \sigma_{J_i}$ for $J_1, J_2 \in \mathcal{J}$. Thus,

$$a\varphi_1 + (1-a)\varphi_2 \geq a\sigma_{J_1} + (1-a)\sigma_{J_2} = \sigma_{aJ_1+(1-a)J_2}.$$

Since $\mathcal{J}$ is convex, $aJ_1 + (1-a)J_2 \in \mathcal{J}$, so $a\varphi_1 + (1-a)\varphi_2 \in C_{\mathcal{J}}$. □

With these claims, we now continue the proof of Theorem 2. Claim 2 implies we can express $C_{\mathcal{J}}$ as the intersection of a countable collection of closed half-spaces that support it (Corollary 7.49 of AB). In other words,

$$C_{\mathcal{J}} = \bigcap_i \{\varphi \in C(W) : \varphi \cdot r_i \geq a_i\}. \tag{3}$$

Moreover, for every i, there exists some $\varphi \in C_{\mathcal{J}}$ such that $\varphi \cdot r_i = a_i$. Consider some closed $E \subset W$ and note that by Urysohn's lemma, we can find a sequence of $\varphi_k \in C(W)$ such that $\varphi_k \to \varphi + 1_E$ and $\varphi_k \geq \varphi$. Since $\varphi_k \geq \varphi \in C_{\mathcal{J}}$, $\varphi_k \in C_{\mathcal{J}}$ by the definition of $C_{\mathcal{J}}$. Thus, $\varphi_k \cdot r_i \geq a_i$ from (3). Since W is compact, it follows from dominated convergence that

$$a_i \leq \lim_k \varphi_k \cdot r_i = (\varphi + 1_E) \cdot r_i = \varphi \cdot r_i + r_i(E) = a_i + r_i(E).$$

Thus, $r_i(E) \geq 0$ for all closed E. Since r_i is a regular Borel measure, we can assume r_i is a probability measure without loss of generality. Finally, note that by the definition of $C_{\mathcal{J}}$, there must exist some $J \in \mathcal{J}$ such that $\varphi \geq \sigma_J$. Since $\sigma_J \geq 0$,

$$a_i = \varphi \cdot r_i \geq \sigma_J \cdot r_i \geq 0.$$

Define the mapping $\psi : H \to W$ such that $\psi(h) = \mathbf{1} - u \circ h$ and let

$$p_i = \frac{1}{2}(r_i \circ \psi) + \frac{1}{2}\delta_{g_i},$$

where $g_i = a_i \bar{x} + (1-a_i)\underline{x}$. Thus,

$$\begin{aligned} U_K(p_i) &= \frac{1}{2}\int_H u_K(h)\, dp_i + \frac{1}{2}a_i = \frac{1}{2}\left(\int_W \min_{\pi \in K} \pi \cdot (1-w)\, dr_i\right) + \frac{1}{2}a_i \\ &= \frac{1}{2}\int_W \left(1 - \max_{\pi \in K} \pi \cdot w\right) dr_i + \frac{1}{2}a_i = \frac{1}{2}(1 - \sigma_K \cdot r_i) + \frac{1}{2}a_i. \end{aligned}$$

Let $f = \frac{1}{2}\bar{x} + \frac{1}{2}\underline{x}$, so $u_K(f) \geq U_K(p_i)$ if and only if

$$\frac{1}{2} \geq \frac{1}{2}(1 - \sigma_K \cdot r_i) + \frac{1}{2}a_i$$
$$a_i \leq \sigma_K \cdot r_i.$$

Let $A_k = \{\delta_f, p_1, \ldots, p_k\}$, so by Claim 1 and (3),

$$\mu(\mathcal{J}) = \mu\bigl(\{K \in \mathcal{K} : \sigma_K \in C_{\mathcal{J}}\}\bigr) = \mu\left(\bigcap_i \{K \in \mathcal{K} : \sigma_K \cdot r_i \geq a_i\}\right)$$
$$= \lim_k \mu\left(\bigcap_{i \leq k} \{K \in \mathcal{K} : \sigma_K \cdot r_i \geq a_i\}\right) = \lim_k \rho_{A_k}(\delta_f).$$

Since $\rho_{A_k}(\delta_f) \geq \tau_{A_k}(\delta_f)$, we have $\mu(\mathcal{J}) \geq \nu(\mathcal{J})$ as desired.

Now suppose $\mu \geq_c \nu$. Let $f \in H$ be constant, where $u(f) = 1 - a$ and $A = \{f, p_1, \ldots, p_k\}$. Again, define the mapping $\psi : H \to W$ such that $\psi(h) = \mathbf{1} - u \circ h$. Note that $u(f) \geq U_K(p_i)$ if and only if

$$1 - a \geq \int_H u_K(h)\, dp_i = \int_H \min_{\pi \in K} \pi \cdot (u \circ h)\, dp_i = 1 - \sigma_K \cdot r_i,$$

where $r_i = p_i \circ \psi^{-1}$. Define

$$\mathcal{J} := \bigcap_{i \leq k} \{K \in \mathcal{K} : \sigma_K \cdot r_i \geq a\}$$

so $\rho_A(\delta_f) = \mu(J)$. Note that $\mathcal{J}$ is closed, convex, and increasing, so

$$\rho_A(\delta_f) = \mu(\mathcal{J}) \geq \nu(\mathcal{J}) = \tau_A(\delta_f)$$

as desired.

A.3 *Proofs for Section 4*

A.3.1 *Extremeness* In this section, we demonstrate the relationship between ex post extremeness and mixture aversion. Since this result is useful for later analysis, for this section, we consider an arbitrary compact metric space Z. Let ρ be a stochastic choice on ΔZ. For finite $A \subset \Delta Z$, let $\text{ext}(A)$ denote the extreme points of A. We define extremeness and mixture aversion for ρ exactly as in Section 4. We say ρ is *monotone* if $A \subset B$ implies $\rho_A(p) \geq \rho_B(p)$.

Lemma 1. *Suppose ρ is monotone. Then ρ is mixture-averse if and only if it is extreme.*

Proof. Note that if ρ is extreme, then ρ is mixture-averse trivially. Suppose ρ is monotone and mixture-averse. We show that ρ is extreme by contradiction. Suppose $\rho_A(\text{ext}(A)) < 1$, so there must be some $p \notin \text{ext}(A)$ such that $\rho_A(p) > 0$. Without loss of generality, assume p is not tied with any point in A (see Lemmas A.2 and A.3 from Lu 2016). By monotonicity, we can also assume $A = \text{ext}(A) \cup \{p\}$ without loss. By the Krein–Milman theorem, we can write

$$p = \sum_{i \leq k} \alpha_i q_i$$

for $\sum_{i \leq k} \alpha_i = 1$, $\alpha_i \in (0, 1)$, and $q_i \in \text{ext}(A) \subset A$. Define $r_1 = q_1$ and

$$\lambda_i := \frac{\alpha_i}{\sum_{j \leq i} \alpha_j}.$$

Recursively define

$$r_i := \lambda_i q_i + (1 - \lambda_i) r_{i-1}$$

and note that $p = r_k$. Let $B := A \cup \{r_1, \ldots, r_{k-1}\}$, so by monotonicity,

$$\rho_B\big(\text{ext}(A)\big) \leq \rho_A\big(\text{ext}(A)\big) < 1.$$

Thus, there must be some $i \in \{2, \ldots, k\}$ such that r_i is not tied with anything in $\text{ext}(A)$ and $\rho_B(r_i) > 0$. By monotonicity again,

$$\rho_{\{r_{i-1}, q_i, r_i\}}(r_i) \geq \rho_B(r_i) > 0.$$

Since $r_i = \lambda_i q_i + (1 - \lambda_i) r_{i-1}$ and ρ is mixture-averse, it means that r_i must be tied with q_i or r_{i-1}. Since $q_i \in \text{ext}(A)$, this means that r_i is tied with r_{i-1}, so $\rho_B(r_{i-1}) = \rho_B(r_i) > 0$. By induction, we conclude that r_i is tied with $r_1 = q_1$, which contradicts the fact that r_i is not tied with anything in $\text{ext}(A)$. Thus, ρ must be extreme. □

A.3.2 *Proof of Theorem 5* In this section, we prove a more general version of Theorem 5 by allowing ν to be *regular*, that is, $v(f) = v(g)$ with probability 0 or 1. As a result, we consider a slightly stronger version of quasiconcavity called quasiconcavity*. If ν excludes ties, then quasiconcavity* and quasiconcavity are equivalent.

Definition 13. Distribution ν is *quasiconcave** if almost surely (a.s.) $v(f) \geq v(g)$ implies $v(af + (1 - a)g) \geq v(g)$, and with strictness if $v(f) > v(g)$ and $a \in (0, 1)$.

Theorem 6. *Let ρ be represented by a regular ν.*

(i) Distribution ν is quasiconcave if and only if ρ is ex post mixture-loving if and only if ρ is ex post hedging.*

(ii) Distribution ν is quasiconvex if and only if ρ is ex post mixture-averse if and only if ρ is ex post extreme.

We start by proving (i). If ρ is ex post hedging, then it is clearly ex post mixture-loving. Suppose ρ is ex post mixture-loving. Let $h = af + (1-a)g$ and $F = \{f, h, g\}$. Note that

$$\begin{aligned}\rho(\delta_f, \delta_h) &= \nu\big(\{v : v(f) \geq v(h)\}\big) \\ &= \nu\big(\{v : v(f) \geq v(h), v(f) \geq v(g)\}\big) + \nu\big(\{v : v(f) \geq v(h), v(f) < v(g)\}\big) \\ &= \rho_{\delta_F}(\delta_f) + \nu\big(\{v : v(g) > v(f) \geq v(h)\}\big),\end{aligned}$$

so $v(g) > v(f) \geq v(h)$ with ν-measure zero. By symmetric argument, $v(f) > v(g) \geq v(h)$ with ν-measure zero. Thus, $v(f) \neq v(g)$ and $\min\{v(f), v(g)\} \geq v(h)$ with ν-measure zero. Note that if f and g are not tied, then $\min\{v(f), v(g)\} < v(h)$ a.s. as desired.

Suppose f and g are tied, so $v(f) = v(g)$ a.s. as ν is regular. Without loss of generality, suppose h is not tied with f. Let

$$b^* := \inf\big\{b : bf + (1-b)h \text{ is tied with } f\big\}.$$

Note that by the continuity of v, $f^* := b^* f + (1-b^*)h$ is tied with f. Thus, $b^* > 0$. Consider $b_k \in (0, b^*)$, so $f_k := b_k f + (1-b_k)h$ is not tied with f and, thus, also not tied with g. By the argument above, $v(h) > \min\{v(f_k), v(g)\}$ a.s. Since $f_k \to f^*$ as $b_k \to b^*$, by the continuity of v, we have a.s.

$$v(h) \geq \min\big\{v\big(f^*\big), v(g)\big\} = \min\big\{v(f), v(g)\big\}$$

as desired. This proves that ν is quasiconcave*.

Finally, to see how quasiconcavity* implies ex post hedging, let $af + (1-a)g \in F$ and note that

$$\begin{aligned}\rho_{\delta_F}(\delta_f) &= \nu\big(\{v : v(f) \geq v(h) \text{ for all } h \in F\}\big) \\ &= \rho_{\delta_{F\cup\{g\}}}(\delta_f) + \nu\big(\{v : v(f) \geq v(h) \text{ for all } h \in F \text{ and } v(f) < v(g)\}\big).\end{aligned}$$

Since ν is quasiconcave*, a.s. $v(g) > v(f)$ implies $v(h) > v(f)$, so the second term above must be zero. This proves that ρ is ex post hedging, concluding the proof for (i).

We now prove (ii). Since ρ has a random utility representation, it is monotone, so Lemma 1 implies that ex post mixture aversion and ex post extremeness are equivalent. Suppose ρ is ex post mixture-averse, and let $h = af + (1-a)g$ and $F = \{f, h, g\}$. Note that if h is tied with either f or g, then clearly $v(h) \leq \max\{v(f), v(g)\}$ a.s. Assume h is tied with neither, so

$$0 = \rho_{\delta_F}(\delta_h) = \nu\big(\{v : v(h) \geq v(f) \text{ and } v(h) \geq v(g)\}\big),$$

so $v(h) < \max\{v(f), v(g)\}$ and ν is quasiconvex. Now suppose ν is quasiconvex. Again, if h is tied with either f or g, then clearly $\rho_{\delta_F}(\delta_{\{f,g\}}) = 1$. Assuming h is tied with neither, we have $\rho_{\delta_F}(\delta_h) = 0$, as ν is quasiconvex. This concludes the proof of Theorem 5.

A.4 *Proof of Theorem 4*

In this section, we prove the main representation theorem. We first show sufficiency and assume ρ satisfies Axioms 1–6. The proof consists of a series of steps. First, we show that choice over constant acts can be represented by a deterministic utility.

Step 1. *There exists a nonconstant vNM utility $u : \Delta X \to \mathbb{R}$ such that for all constant δ_F, $\rho_{\delta_F}(\delta_f) = 1$ if $u(f) \geq u(g)$ for all $g \in F$.*

Proof. Define a stochastic choice $\hat{\rho}$ on finite $F \subset H_c$ such that $\hat{\rho}_F = \rho_{\delta_F}$. We show that $\hat{\rho}$ satisfies the random expected utility axioms. If $F \subset G \subset H_c$, then $\delta_F \subset \delta_G$, so $\hat{\rho}_F(f) \geq \hat{\rho}_G(f)$ by Axiom 1.1. Thus, $\hat{\rho}$ satisfies monotonicity. Note that Axiom 3 implies that $\delta_{af+(1-a)g}$ and $a\delta_f \oplus (1-a)\delta_g$ are tied for all $f, g \in H_c$. Hence, Axioms 1.2 and 1.3 imply that $\hat{\rho}$ satisfies linearity and extremeness, respectively. For continuity, note that mapping $F \to \delta_F$ is a homeomorphism (see Theorem 15.8 of AB). Thus, by Axiom 1.4, $\hat{\rho}$ is also continuous. Hence, $\hat{\rho}$ has a random expected utility representation (see Theorem S.1 of Lu 2016). Axiom 4 implies that this random utility must be deterministic. In other words, there is a vNM utility $u : \Delta X \to \mathbb{R}$ such that $\rho_{\delta_F}(\delta_f) = 1$ if $u(f) \geq u(g)$ for all $g \in F \subset H_c$. Axioms 5 and 6 imply that u must be nonconstant. □

Given Step 1, we can now choose $\bar{x}, \underline{x} \in X$ such that $1 = u(\bar{x}) \geq u(x) \geq u(\underline{x}) = 0$ for all $x \in X$. The next step shows that these are indeed the best and worst prizes.

Step 2. *It must be that $\rho(\delta_{\underline{x}}, \delta_{\bar{x}}) = 0$ and $\rho(\delta_{\bar{x}}, \delta_f) = \rho(\delta_f, \delta_{\underline{x}}) = 1$ for all $f \in H$.*

Proof. Note that $\rho(\delta_{\underline{x}}, \delta_{\bar{x}}) = 0$ follows from Step 1 and the definitions of $\bar{x}$ and $\underline{x}$. Moreover, for any $f \in H$ and $s \in S$,

$$\rho(\delta_{\bar{x}}, \delta_{f(s)}) = \rho(\delta_{f(s)}, \delta_{\underline{x}}) = 1.$$

By Axiom 5, $\rho(\delta_{\bar{x}}, \delta_f) = \rho(\delta_f, \delta_{\underline{x}}) = 1$ as desired. □

Recall that $H \subset \mathbb{R}^{S \times X}$ and define the supnorm

$$|f - g| := \sup_{s,x} \big| f(s,x) - g(s,x) \big|.$$

The last step shows that ρ satisfies a Lipschitz continuity condition from Lu and Saito (2020).

Step 3. *For any $a \in [0,1]$, if $|f - g| \leq a/|X|$, then*

$$\rho\big(a\delta_{\bar{x}} + (1-a)\delta_f, a\delta_{\underline{x}} + (1-a)\delta_g\big) = 1.$$

Proof. Let $m = |X|$ and suppose $|f - g| \leq a/m$. Thus, for all $s, x \in S \times X$,

$$g(s,x) - f(s,x) \leq \frac{a}{m}.$$

Hence, for any $s \in S$,

$$u\big(g(s)\big) - u\big(f(s)\big) = \sum_x u(x)\big(g(s, x) - f(s, x)\big) \leq \sum_x u(x)\frac{a}{m} \leq a.$$

As a result,

$$\begin{aligned} u\big(g(s)\big) &\leq a + u\big(f(s)\big) \\ (1-a)u\big(g(s)\big) &\leq a(1-a) + (1-a)u\big(f(s)\big) \\ au(\underline{x}) + (1-a)u\big(g(s)\big) &\leq au(\bar{x}) + (1-a)u\big(f(s)\big). \end{aligned}$$

Let $f' := a\bar{x} + (1-a)f$ and $g' := a\underline{x} + (1-\alpha)g$, so $u(f'(s)) \geq u(g'(s))$ for all $s \in S$. By Step 1, $\rho(\delta_{f'(s)}, \delta_{g'(s)}) = 1$ for all $s \in S$. Hence, by Axiom 5,

$$1 = \rho(\delta_{f'}, \delta_{g'}) = \rho(\delta_{a\bar{x}+(1-a)f}, \delta_{a\underline{x}+(1-\alpha)g}).$$

By Axiom 3, $\delta_{a\bar{x}+(1-a)f}$ is tied with $a\delta_{\bar{x}} \oplus (1-a)\delta_f$ and $\delta_{a\underline{x}+(1-\alpha)g}$ is tied with $a\delta_{\underline{x}} \oplus (1-a)\delta_g$. Hence,

$$1 = \rho\big(a\delta_{\bar{x}} \oplus (1-a)\delta_f, a\delta_{\underline{x}} \oplus (1-a)\delta_g\big)$$

as desired. □

Axioms 1.1–1.4 and Steps 2 and 3 imply that ρ satisfies the sufficient conditions for a random expected utility representation (see Theorem 5 of Lu and Saito 2020). Thus, there exists a distribution ν on Lipschitz continuous utilities $v : H \to [0, 1]$ such that $v(\bar{x}) = 1$, $v(\underline{x}) = 0$, and

$$\rho_A(p) = \nu\big(\{v : V(p) \geq V(q) \text{ for all } q \in A\}\big),$$

where $V(p) := \int_H v(f)\,dp$. Moreover, ν is regular in that $V(p) = V(q)$ with probability 0 or 1.

Finally, we show that v satisfies the axioms of maxmin expected utility a.s. Note that Axiom 2 implies that v is quasiconcave by Theorem 6. The c-independence and state-by-state monotonicity follow from Axioms 3 and 5, respectively. Note that v is continuous and $v(\bar{x}) > v(\underline{x})$. Thus, by Theorem 1 of Gilboa and Schmeidler (1989), we can write

$$v(f) = u_K(f) = \min_{\pi \in K} \pi \cdot (u \circ f).$$

This establishes the sufficiency of the axioms.

We now establish necessity. Note that Axioms 1.1–1.4 follow from the characterization of random expected utility (see Theorem S.1 of Lu 2016). Since $u_K(\cdot)$ is concave in ex post mixing and any concave function is quasiconcave*, Axiom 2 follows from Theorem 6. To see Axiom 3, note that for constant h,

$$u_K\big(af + (1-a)h\big) = au_K(f) + (1-a)u_K(h).$$

Finally, Axioms 4–6 follow immediately from the representation. This concludes the proof.

A.5 *Identification without ex ante randomization*

In this section, we show that without ex ante randomization, identification is possible with more than just binary menus.

Theorem 7. *Let ρ and τ be represented by (μ, u) and (ν, v), respectively. Then the following statements are equivalent:*

(i) We have $\rho_{\delta_F}(\delta_f) = \tau_{\delta_F}(\delta_f)$ for all constant $f \in H$ and $\delta_f \in \delta_F$.

(ii) We have $\mu = \nu$ and $u = \alpha v + \beta$ for $\alpha > 0$.

Proof. Note that (ii) implying (i) is straightforward, so suppose (i) is true. By the same argument as the proof for Theorem 1, we can assume $u = v$ without loss of generality, and let $1 = u(\bar{x}) \geq u(x) \geq u(\underline{x}) = 0$ for $\bar{x}, \underline{x} \in X$ and all $x \in X$.

Let $W := [0, 1]^n$, where $n = |S|$. Now, for any $w_i \in W$ and $a_i \in [0, 1]$, we can find some $g_i \in H$ such that

$$\mathbf{1} - u \circ g_i = \frac{1}{2} w_i + \frac{1}{2}(1 - a_i)\mathbf{1} \in W.$$

Let $f = \frac{1}{2}\bar{x} + \frac{1}{2}\underline{x}$ be a constant act and note that $u_K(f) \geq u_K(g_i)$ if and only if

$$\begin{aligned} 1 - \sigma_K(\mathbf{1} - u \circ g_i) &\leq \frac{1}{2} \\ \sigma_K\left(\frac{1}{2} w_i + \frac{1}{2}(1 - a_i)\mathbf{1}\right) &\geq \frac{1}{2} \\ \sigma_K(w_i) &\geq a_i. \end{aligned}$$

Define

$$\mathcal{J}_i := \{K \in \mathcal{K} : \sigma_K(w_i) \geq a_i\}.$$

Given a finite set I, let $F = \{g_i : i \in I\} \cup \{f\}$, so

$$\mu\left(\bigcap_{i \in I} \mathcal{J}_i\right) = \rho_{\delta_F}(\delta_f) = \tau_{\delta_F}(\delta_f) = \nu\left(\bigcap_{i \in I} \mathcal{J}_i\right). \tag{4}$$

Now

$$\begin{aligned} \mu\left(\bigcap_{i=1}^m \mathcal{J}_i^c\right) &= \mu\left(\left(\bigcup_{i=1}^m \mathcal{J}_i\right)^c\right) = 1 - \mu\left(\bigcup_{i=1}^m \mathcal{J}_i\right) \\ &= 1 - \sum_{j=1}^m (-1)^{j-1} \sum_{I \subset \{1,\ldots,m\}, |I| = j} \mu\left(\bigcap_{k \in I} \mathcal{J}_k\right), \end{aligned} \tag{5}$$

where the last equation follows from the inclusion–exclusion principle. Now, given any $J \in \mathcal{K}$, we can find a sequence $a_{jk} \searrow \sigma_K(w_j)$ for any w_j in some dense set of W. This means we can also find a sequence $\mathcal{J}_i$ such that

$$\bigcap_i \mathcal{J}_i^c \searrow \{K \in \mathcal{K} : \sigma_K(w) \leq \sigma_J(w) \text{ for all } w \in W\},$$

so by continuity,

$$\mu\left(\bigcap_i \mathcal{J}_i^c\right) \to \mu(\{K \in \mathcal{K} : K \subset J\}).$$

Together with (4) and (5), this implies that for all $J \in \mathcal{K}$,

$$\mu(\{K \in \mathcal{K} : K \subset J\}) = \nu(\{K \in \mathcal{K} : K \subset J\}).$$

Since containment functionals completely characterize the distribution of convex compact sets (Theorem 7.8 of Molchanov 2005), $\mu = \nu$ as desired. □

References

Abdellaoui, Mohammed, Aurelien Baillon, Laetitia Placido, and Peter Wakker (2011), "The rich domain of uncertainty, source functions and their experimental implementation." *American Economic Review*, 101, 695–723. [539]

Agranov, Marina and Pietro Ortoleva (2017), "Stochastic choice and preferences for randomization." *Journal of Political Economy*, 125, 40–68. [542]

Ahn, David, Syngjoo Choi, Douglas Gale, and Shachar Kariv (2014), "Estimating ambiguity aversion in a portfolio choice experiment." *Quantitative Economics*, 5, 195–223. [539]

Ahn, David S. and Todd Sarver (2013), "Preference for flexibility and random choice." *Econometrica*, 81, 341–361. [542]

Aliprantis, Charalambos D. and Kim C. Border (2006), *Infinite Dimensional Analysis*. Springer, Berlin. [559]

Anscombe, Francis J. and Robert J. Aumann (1963), "A definition of subjective probability." *Annals of Mathematical Statistics*, 34, 199–205. [540, 543]

Apesteguia, Jose, Miguel Ballester, and Jay Lu (2017), "Single-crossing random utility models." *Econometrica*, 85, 661–674. [556]

Aumann, Robert (1965), "Integrals of set-valued functions." *Journal of Mathematical Analysis and Applications*, 12, 1–12. [550]

Becker, Gordon, Morris DeGroot, and Jacob Marschak (1963), "An experimental study of some stochastic models for wagers." *Behavioral Science*, 8, 199–202. [542]

Block, Henry David and Jacob Marschak (1960), "Random orderings and stochastic theories of response." In *Contributions to Probability and Statistics* (Ingram Olkin, ed.), 97–132, Stanford University Press. [542]

Bose, Subir, Emre Ozdenoren, and Andreas Pape (2006), "Optimal auctions with ambiguity." *Theoretical Economics*, 1, 411–438. [543]

Caballero, Ricardo J. and Arvind Krishnamurthy (2008), "Collective risk management in a flight to quality episode." *Journal of Finance*, 63, 2195–2230. [539, 543]

Cerreia-Vioglio, Simone, David Dillenberger, Pietro Ortoleva, and Gil Riella (2019), "Deliberately stochastic." *American Economic Review*, 109, 2425–2445. [544]

Debreu, Gerard (1967), "Integration of correspondences." In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 4, 351–372, University of California Press. [551]

Dekel, Eddie (1986), "An axiomatic characterization of preferences under uncertainty: Weakening the independence axiom." *Journal of Economic Theory*, 40, 304–318. [557]

Dow, James and Sérgio Ribeiro da Costa Werlang (1992), "Uncertainty aversion, risk aversion, and the optimal choice of portfolio." *Econometrica*, 60, 197–204. [543]

Duraj, Jetlir (2018), "Dynamic random subjective expected utility." Unpublished paper, arXiv1808.00296. [542]

Easley, David and Maureen O'Hara (2009), "Ambiguity and nonparticipation: The role of regulation." *Review of Financial Studies*, 22, 1817–1843. [540, 543]

Ellsberg, Daniel (1961), "Risk, ambiguity, and the Savage axioms." *Quarterly Journal of Economics*, 75, 643–669. [546]

Epstein, Larry G. and Igor Kopylov (2007), "Cold feet." *Theoretical Economics*, 2, 231–259. [542]

Epstein, Larry G. and Martin Schneider (2003), "Recursive multiple-priors." *Journal of Economic Theory*, 113, 1–31. [542, 543, 552]

Epstein, Larry G. and Martin Schneider (2010), "Ambiguity and asset markets." *Annual Review of Financial Economics*, 2, 315–346. [543]

Epstein, Larry G. and Tan Wang (1994), "Intertemporal asset pricing under Knightian uncertainty." *Econometrica*, 62, 283–322. [543]

Falmagne, Jean-Claude (1978), "A representation theorem for finite random scale systems." *Journal of Mathematical Psychology*, 18, 52–72. [541, 542]

Frick, Mira, Ryota Iijima, and Tomasz Strzalecki (2019), "Dynamic random utility." *Econometrica*, 87, 1941–2002. [542]

Fudenberg, Drew, Ryota Iijima, and Tomasz Strzalecki (2015), "Stochastic choice and revealed perturbed utility." *Econometrica*, 83, 2371–2409. [542]

Fudenberg, Drew and Tomasz Strzalecki (2015), "Dynamic logit with choice aversion." *Econometrica*, 83, 651–691. [542]

Gajdos, Thibault, Takashi Hayashi, Jean-Marc Tallon, and Jean-Christophe. Vergnaud (2008), "Attitude toward imprecise information." *Journal of Economic Theory*, 140, 27–65. [546]

Ghirardato, Paolo, Fabio Maccheroni, and Massimo Marinacci (2004), "Differentiating ambiguity and ambiguity attitude." *Journal of Economic Theory*, 118, 133–173. [545]

Ghirardato, Paolo and Massimo Marinacci (2002), "Ambiguity made precise: A comparative foundation." *Journal of Economic Theory*, 102, 251–289. [549, 550]

Gilboa, Itzhak and David Schmeidler (1989), "Maxmin expected utility with non-unique prior." *Journal of Mathematical Economics*, 18, 141–153. [539, 540, 544, 554, 565]

Gilboa, Itzhak and David Schmeidler (1993), "Updating ambiguous beliefs." *Journal of Economic Theory*, 59, 33–49. [542]

Gul, Faruk and Wolfgang Pesendorfer (2006), "Random expected utility." *Econometrica*, 74, 121–146. [541, 542, 553, 555, 557]

Hanany, Eran and Peter Klibanoff (2007), "Updating preferences with multiple priors." *Theoretical Economics*, 2, 261–298. [543]

Kahneman, Daniel and Amos Tversky (1979), "Prospect theory: An analysis of decision under risk." *Econometrica*, 47, 263–291. [542]

Knight, Frank Hyneman (1921), *Risk, Uncertainty and Profit*. Houghton Mifflin. [539]

Kopylov, Igor (2009), "Choice deferral and ambiguity aversion." *Theoretical Economics*, 4, 199–225. [546]

Lin, Yi-Hsuan (2019), "Stochastic choice and rational inattention." Unpublished paper, Institute of Economics, Academia Sinica. [542]

Lin, Yi-Hsuan (2020), "Random non-expected utility: Non-uniqueness." Unpublished paper, arXiv:2009.04173. [541, 542, 548, 557]

Lu, Jay (2016), "Random choice and private information." *Econometrica*, 84, 1983–2027. [541, 542, 544, 546, 548, 554, 562, 564, 565]

Lu, Jay and Kota Saito (2018), "Random intertemporal choice." *Journal of Economic Theory*, 177, 780–815. [542, 548]

Lu, Jay and Kota Saito (2020), "Repeated choice: A theory of stochastic intertemporal preferences." Working paper, Social Science Working Paper, 1449. California Institute of Technology. [546, 548, 555, 564, 565]

McFadden, Daniel and Marcel K. Richter (1990), "Stochastic rationality and revealed stochastic preference." In *Preferences, Uncertainty and Optimality: Essays in Honor of Leonid Hurwicz* (John S. Chipman, Daniel McFadden, and Marcel K. Richter, eds.), 161–186, Westview Press, Boulder, Colorado. [542]

Molchanov, Ilya (2005), *Theory of Random Sets*. Springer, London. [548, 551, 552, 559, 567]

Raiffa, Howard (1961), "Risk, ambiguity, and the Savage axioms: Comment." *Quarterly Journal of Economics*, 75, 690–694. [544]

Ryan, Matthew (2018), "Uncertainty and binary stochastic choice." *Economic Theory*, 65, 629–662. [542]

Saito, Kota (2015), "Preferences for flexibility and randomization under uncertainty." *American Economic Review*, 105, 1246–1271. [542, 544]

Sarin, Rakesh and Peter P. Wakker (1998), "Dynamic choice and nonexpected utility." *Journal of Risk and Uncertainty*, 17, 87–119. [542]

Seo, Kyoungwon (2009), "Ambiguity and second-order belief." *Econometrica*, 77, 1575–1605. [540, 542, 543, 545]

Siniscalchi, Marciano (2011), "Dynamic choice under ambiguity." *Theoretical Economics*, 6, 379–421. [543]

Co-editor Ran Spiegler handled this manuscript.

Manuscript received 6 June, 2019; final version accepted 25 June, 2020; available online 28 July, 2020.